REPORT RESUMES

ED 003 374 24
RESEARCH IN WRITTEN COMPOSITION.
BY- BRADDOCK, RICHARD AND OTHERS
NATIONAL COUNCIL OF TEACHERS OF ENGLISH
REPORT NUMBER CRP-E-004 PUB DATE 63
IOWA UNIV., IOWA CITY
EDRS PRICE MF-$0.27 HC-$5.96 149P.

DESCRIPTORS- *GRAMMAR, *COMPOSITION (LITERARY), PARAGRAPH COMPOSITION, *WRITING, COMPOSITION SKILLS (LITERARY), *TEACHING METHODS, EDUCATIONAL RESEARCH, LEARNING PROCESSES, RESOURCE GUIDES, IOWA CITY, IOWA, CHAMPAIGN, ILLINOIS

THIS IS THE REPORT OF A COMMITTEE WHICH IDENTIFIED FIVE STUDIES WHICH IN ITS OPINION WERE SOUNDLY BASED UPON CONTROLLED EXPERIMENTATION AND TEXTUAL ANALYSIS. THE FIVE STUDIES ARE BRIEFLY SUMMARIZED AND THE IMPLICATIONS OF EACH FOR THE IMPROVEMENT OF TEACHING ENGLISH COMPOSITION ARE PRESENTED. (JK)

ED003374

REFERENCE
COPY

COOPERATIVE
RESEARCH
PROGRAM

No. E-004

RCH IN WRITTEN COMPOSITION

U. S. Office of Education
Dept. of Health, Education, & Welfare
Washington D.C. 20202

By

Richard Braddock,
Richard Lloyd-Jones
and
Lowell Schoer

NATIONAL COUNCIL OF TEACHERS OF ENGLISH
508 South Sixth Street Champaign, Illinois

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

RESEARCH IN WRITTEN COMPOSITION

By

Richard Braddock
Richard Lloyd-Jones
and
Lowell Schoer
all of the
University of Iowa

Under the supervision and with the assistance of the

NCTE COMMITTEE ON THE STATE OF KNOWLEDGE ABOUT COMPOSITION

Alvina Treut Burrows, *New York University*
Richard Corbin, *Hunter College High School*
Mary Elizabeth Fowler, *Central Connecticut State College*
Dora V. Smith, *University of Minnesota*
Erwin R. Steinberg, *Carnegie Institute of Technology*
Priscilla Tyler, *University of Illinois*
Harold B. Allen, *University of Minnesota, ex officio*
James R. Squire, *NCTE, ex officio*

Chairman: Richard Braddock, *University of Iowa*
Associate Chairman: Joseph W. Miller, *Moorhead State College*

Supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare

NATIONAL COUNCIL OF TEACHERS OF ENGLISH
508 South Sixth Street Champaign, Illinois
1963

ED 003374

COMMITTEE ON PUBLICATIONS
of the
NATIONAL COUNCIL OF TEACHERS OF ENGLISH

JAMES R. SQUIRE, *NCTE Executive Secretary, Chairman*
JARVIS E. BUSH, *Wisconsin State College, Oshkosh*
AUTREY NELL WILEY, *Texas Woman's University*
MIRIAM E. WILT, *Temple University*
ENID M. OLSON, *NCTE Director of Publications*

Copyright 1963
National Council of Teachers of English

TABLE OF CONTENTS

I. The Preparation of This Report 1

II. Suggested Methods of Research 5

- Rating Compositions 6
  - The writer variable 6
  - The assignment variable: the topic—the mode of discourse —the time afforded for writing—the examination situation 7
  - The rater variable: personal feelings—rater fatigue . . . 10
  - The colleague variable: a common set of criteria—practice rating 11
- Frequency Counts 15
  - Clarifying examples for each type of item 16
  - Standard classification of types of items 16
  - Control or sampling of compositions according to topic, mode of discourse, and writer characteristics 17
  - Need for analyses of rhetorical constructions 18
  - Need for imaginative approaches to frequency counts . . 18
  - Counting types of responses by various kinds of writers to various types of situations 19
  - Reporting frequency per hundred or thousand words . . 20
  - Using the cumulative-average technique of sampling . . 20
  - Focusing investigation on narrower, more clearly defined areas and exploring them more thoroughly and carefully 21
  - Seeking key situations which are indices of larger areas of concern 21
- General Considerations 21
  - Attitude of the investigator 22
  - Meaning of terms and measures: clarity of terms and measures—direct observation—validity of assumptions —reliability of criterion application 23
  - Planning of procedures: planning before initiating research —using appropriate and consistent statistical procedures 24
  - Controlling of variables: selection of teachers and students —control of "outside influences"—control of additional influences 25
- Need for trials and checks 26

Reporting of results: complete enough to permit replication—limitation of conclusions to type of population investigated—inclusion of raw data—use of standard methods of description and statistical analysis—allowing for the microfilm medium 27

III. The State of Knowledge about Composition 29

Environmental Factors Influencing Composition 29
  Primacy of the writer's experiences 29
  Influence of socioeconomic background 30
  Composition interests 30
  Flow of words 31
  Need for case studies 31
  Need for longitudinal studies 32
Instructional Factors Influencing Composition 33
  Student correction 34
  Frequency of writing 34
  Student revision 35
  Nature of marking and grading 36
  Ineffectiveness of instruction in formal grammar . . . 37
Rhetorical Considerations 38
  Distinctive tendencies of good writers 39
  Organizational factors 39
  Effects on readers 39
Objective Tests versus Actual Writing as Measures of Writing 40
  Interlinear tests 40
  "Self-evident" invalidity of objective tests 41
  Unreliable grading of compositions 41
  Reliable grading of compositions 41
  More on invalidity of objective tests 42
  Reliability of objective tests 43
  Varying emphases in college instruction 43
  Use of objective tests for rough sorting of many students 44
  Basing diagnosis of individual needs on actual writing . . 45
  Evaluating writing from several compositions 45
Other Considerations 45
  Size of English classes 45

Lay readers 46
Teaching by television 47
Writing vocabulary 48
Spelling 49
Handwriting 50
Typewriting 51
Relationships of oral and written composition 51
Unexplored territory 52

IV. Summaries of Selected Research 55

Basis for Selecting These Studies 55
Explanation of Statistical Terms 56
The Buxton Study 58
The Harris Study 70
The Kincaid Study 83
The Smith Study 95
The Becker Study 107

V. References for Further Research 117

Summaries and Bibliographies 117
Indices and Abstracts 118
Bibliography for This Study 118

I.

THE PREPARATION OF THIS REPORT

Reading a report, like driving over a bridge, is an act of faith—faith that the other fellow has done his job well. The writers of this pamphlet do not ask that the reader's faith be blind. To permit him to evaluate their work, they explain in this chapter the procedures resulting in their generalizations. The explanation also provides an opportunity to acknowledge the assistance rendered by colleagues throughout the United States and in Canada and England.

The impetus to prepare this report came from the Executive Committee of the National Council of Teachers of English. Concerned over the nature of public pronouncements about how writing should be taught—the sound and the wild seem to share space equally in the press—the Executive Committee appointed an ad hoc Committee on the State of Knowledge about Composition "to review what is known and what is not known about the teaching and learning of composition and the conditions under which it is taught, for the purpose of preparing for publication a special scientifically based report on what is known in this area." The membership of the ad hoc committee is named on the title page.

In April, 1961, the committee met in Washington to clarify the purposes of its task and to plan procedures. It agreed, among other things, to limit its task to written composition and, more particularly, to studies in which some actual writing was involved (not studies entirely restricted to objective testing and questionnaires). The committee further decided to use only research employing "scientific methods," like controlled experimentation and textual analysis. At the suggestion of the Executive Committee, the ad hoc committee set as its goal the identification of the dozen or so most soundly based studies of the foregoing type. (Actually, the committee finally identified five such studies, each of which is summarized in detail in Chapter IV.)

First instructed to complete the manuscript in six to eight months, the ad hoc committee soon realized that a review of "all" the research on composition was a prodigious undertaking which would necessitate a much longer period of preparation. Consequently, as it began its task, the chairman of the committee applied to the Office of Education, U. S. Department of Health, Education, and Welfare, for a Cooperative Research Program grant. A grant was awarded in the amount of $13,345, supplemented by an allocation of $4,397 from the University of Iowa.

Before the grant was approved, the ad hoc committee had surveyed some 20 summaries and bibliographies (*Dissertation Abstracts, Psychological Abstracts, Review of Educational Research*, etc.) for titles of studies which seemed pertinent. From more than 1,000 bibliographic citations discovered by the committee, enough apparently tangential references were eliminated to reduce the number to 485 items, which were typed in a dittoed list late in the summer of 1961. The problem then was to screen the studies to determine which should be read carefully.

Because about half of the 485 studies were unpublished, the assistance of colleagues on other campuses was requested. Whenever three or more dissertations from a single campus were on the list, the services of a colleague on that campus were solicited to read the studies and advise the committee on whether or not to study them more carefully. The following people helped in this fashion:

Richard S. Beal, Boston University
Margaret D. Blickle, The Ohio State University
Francis Christensen, University of Southern California
Robert W. DeLancey, Syracuse University
Wallace W. Douglas, Northwestern University
David Dykstra, University of Kansas
Margaret Early, Syracuse University (then visiting Teachers College, Columbia University)
William H. Evans, University of Illinois
Donald J. Gray, Indiana University
Catherine Ham, University of Chicago
Arnold Lazarus, Purdue University (then University of Texas)
V. E. Leichty, Michigan State University
William McColly, University of Wisconsin
John C. McLaughlin, University of Iowa
George E. Murphy, The Pennsylvania State University
Leo P. Ruth, University of California, Berkeley
George S. Wykoff, Purdue University

The large majority of the 485 studies remained, of course, and these were apportioned among the members of the ad hoc committee to screen. To encourage careful screening, each person was requested to fill out a three-page questionnaire for each study he recommended.

Between the number of manuscripts recommended and the number so far inaccessible because of location on other campuses (some of them mimeographed reports not in libraries) several hundred items were still to be read. It was at this point, in the spring of 1962, that funds from the Office of Education and University of Iowa became available, providing the time and money needed to order unpublished material through interlibrary loan and to purchase microfilms, to draw together the findings and to write the pamphlet. Under the provisions of the Office of Education grant, the main responsibility for the project had to be focused in one university. Consequently, a director and two associate directors on the University of Iowa faculty were released from some of their ordinary responsibilities to accomplish these tasks—Richard Braddock, associate professor of English and Rhetoric; Richard Lloyd-Jones, associate professor of English; and Lowell Schoer, assistant professor of Educational Psychology. The grant made it possible to obtain the services of two special consultants—Alvina Treut Burrows, consultant in Elementary Education; and Porter G. Perrin, consultant in Rhetoric, who died before his invaluable experience could be utilized.

By the end of the summer, 1962, it was possible to construct a list of studies which so far had passed the screening procedures. The directors had not had time to rescreen all recommended studies, and some items were added to the list which no one had yet examined. This list of some 100 studies was submitted to research specialists with a request for additional titles which might have been overlooked or perhaps too hastily screened. The following specialists suggested over fifty new titles to consider as well as some mimeographed bibliographies which the directors did not systematically screen:

Paul B. Diederich, Educational Testing Service
Carl J. Freudenreich, New York State Education Department
Robert M. Gorrell, University of Nevada
S. I. Hayakawa, Editor, *Etc.*
Ernest Horn, University of Iowa
Arno Jewett, U. S. Office of Education
Walter V. Kaulfers, University of Illinois
Albert R. Kitzhaber, University of Oregon

Lou LaBrant, Dillard University
Walter Loban, University of California, Berkeley
Helen K. Mackintosh, U. S. Office of Education
Joseph Mersand, Jamaica High School
Edwin L. Peterson, University of Pittsburgh
Robert C. Pooley, University of Wisconsin
C. B. Routley, Canadian Education Association
David H. Russell, University of California, Berkeley
Ruth Strickland, Indiana University
Stephen Wiseman, University of Manchester

In addition, a number of other people volunteered suggestions or sent material, including Mary Long Burke, Harvard University; Ruth Godwin, University of Alberta; Robert Hogan, NCTE; Elsie L. Leffingwell, Carnegie Institute of Technology; and Harold C. Martin, Harvard University.

Each of the three directors now proceeded to reread each of the studies which had been recommended so far, noting the strengths and weaknesses as a basis for periodic conferences in which they discussed six or eight studies in an hour. At these conferences they also decided which research to recommend to the ad hoc committee for the highly selected studies to be summarized at length in the final report.

During the Christmas vacation, 1962, the three directors and the members of the ad hoc committee met to discuss the selected studies and the nature of the final report. Many problems were discussed and suggestions made to guide the directors. After that meeting, the directors completed their reading and discussion of the studies and wrote the report.

Several steps were taken to check the accuracy of this report. The summaries of the five selected studies were submitted to the authors of the original research to insure that the summaries and interpretative parenthetical comments were accurate. Copies of the report were also emended by the members of the ad hoc committee and by the Committee on Publications of the National Council of Teachers of English.

Special acknowledgments are extended to the following consulting readers, who offered helpful suggestions in the final preparation of the manuscript: Margaret J. Early, Syracuse University; Arno Jewett, U. S. Office of Education; Albert R. Kitzhaber, University of Oregon; and David H. Russell, University of California, Berkeley.

II.

SUGGESTED METHODS OF RESEARCH

Hearing about the project of which this report is the result, a colleague wrote, "What is the sense of attempting an elaborate empirical study if there is no chance of controlling the major elements in it? I think . . . that the further we get away from the particularities of the sentence, the less stable our 'research' becomes. I do not for that reason think there should be no study and speculation about the conditions for teaching composition and about articulation, grading, and the like, but I do think that it is something close to a mockery to organize these structures as though we were conducting a controlled experiment."

Certainly there is much truth in that statement, especially if one takes it as a comment on the bulk of the research which has been conducted thus far on the teaching of written composition. But research in this area, complex though it may be (especially when it deals with the "larger elements" of composition, not merely with grammar and mechanics), has not frequently been conducted with the knowledge and care that one associates with the physical sciences. Today's research in composition, taken as a whole, may be compared to chemical research as it emerged from the period of alchemy: some terms are being defined usefully, a number of procedures are being refined, but the field as a whole is laced with dreams, prejudices, and makeshift operations. Not enough investigators are really informing themselves about the procedures and results of previous research before embarking on their own. Too few of them conduct pilot experiments and validate their measuring instruments before undertaking an investigation. Too many seem to be bent more on obtaining an advanced degree or another publication than on making a genuine contribution to knowledge, and a fair measure of the blame goes to the faculty adviser or journal editor who permits or publishes such irresponsible work. And far too few of those who have conducted an initial piece of research follow it with further exploration or replicate the investigations of others.

Composition research, then, is not highly developed. If researchers wish to give it strength and depth, they must reexamine critically the structure and techniques of their studies. To that end, this report now surveys some of the methods and elements of design in composition research. The hope is that serious investigators will find them useful in advancing the research in composition. An intention is also to reveal the considerations used in selecting the five "most soundly based" studies summarized at length in Chapter IV.

Rating Compositions

The Writer Variable

One of the fundamental measures in research into the teaching of composition is, of course, the general evaluation of actual writing. Often referred to as measures of writing *ability,* composition examinations are always measures of writing *performance;* that is, when one evaluates an example of a student's writing, he cannot be sure that the student is fully using his ability, is writing as well as he can. Something may be causing the student to write below his capacity: a case of the sniffles, a gasoline lawnmower outside the examination room, or some distracting personal concern. If a student's writing performance is consistently low, one may say that he has *demonstrated* poor ability, but often one cannot say positively that he *has* poor ability; perhaps the student has latent writing powers which can be evoked by the right instruction, the appropriate topic, or a genuine need for effective writing in the student's own life. It is not difficult to see why Kincaid discovered, as reported in Chapter IV, that, at least with college freshmen, *the day-to-day writing performance of individuals varies, especially the performance of better writers.*[1] Similarly, C. C. Anderson found that 71 percent of the 55 eighth grade students he examined on eight different occasions "showed evidence of composition fluctuation" apart from the discrepancies attributable to the raters.[2] These and other studies point clearly to the existence of a *writer variable* which must be taken into account when rating compositions for research purposes.

Although it is obvious that the writer variable cannot be controlled, certainly allowances should be made for it. If it is desirable to evaluate a student's composition when it is as good as his performance typically gets, he should write at least twice, once on each of at least two different occasions, the rating of the better paper being used as the measure of his writing performance.[3] Some investigators have maintained that variations in the day-to-day writing performance of individual students "cancel each other out" when the mean rating of a large group of students is considered. But this assumption is false if Kincaid's finding is true that the performance of good writers varies more than the performance of poor writers; the mean rating of the single papers from each of the good writers would not reflect their typically good writing as closely as the mean rating of single papers from poor writers would reflect their typically good writing. The importance of this realization is emphasized by the fact that annual increments in the level of writing performance have usually been reported as small—as approximately one point on a rating scale reaching from 1 to 20, or as 5 percent. Especially, then, if an investigator wishes to measure individual students' improvement in writing, he should provide for at least two writing occasions as a pretest, at least two as a post-test, and count the rating only of the better composition on each occasion. If three writing occasions are used for each test, it may be wisest to average the ratings of the two best papers, but more research needs to be done on this possibility.[4]

---

[1] Gerald L. Kincaid, "Some Factors Affecting Variations in the Quality of Students' Writing" (Unpublished Ed.D. dissertation, [Michigan State College] Michigan State University, 1953).

[2] C. C. Anderson, "The New STEP Essay Test as a Measure of Composition Ability," *Educational and Psychological Measurement,* XX (Spring, 1960), 95-102.

The Assignment Variable

A second variable—one which can be controlled but often is not—is the assignment variable, with its four aspects: *the topic, the mode of discourse, the time afforded for writing, and the examination situation.*

Significant variations in the writing performance of eleven-year-olds who wrote on different *topics* four months apart have been discovered, for example, by Wiseman and Wrigley.[5] The children had a choice from the same set of five topics each time, but the second time had to select a topic different from the one they wrote on the first time. Evidently the investigators assumed that variations in quality of writings were associated with variations in topics not because of the topics themselves but because of the writers' abilities or the raters' idiosyncrasies. Although Wiseman and Wrigley attributed "the bulk of differences in title means [average rating for all papers written on the same topic, or title] . . . to the ability of the children rather than to the idiosyncrasies of the markers," only four raters were involved and it cannot be determined how representative they were of raters in general. Until more conclusive research has been conducted, it seems safest to select topics with care when rating compositions for purposes of research. Wiseman and Wrigley concluded that examinees might as well be given a choice of topics; the practice of the College Entrance Examination Board suggests that a single topic should be used, controlling the effects of the topic on the quality of the writing. But, whichever practice is correct, it seems very advisable when using compositions as pretests and post-tests to consider carefully the abstractness of the topics and their familiarity to the entire group of examinees. In planning composition examinations for students from a wide range of backgrounds, it seems especially necessary to consider the students' variations in intellectual maturity, knowledge, and socioeconomic background. The national examiner is not adequately controlling the topic who blithely assigns the single subject "My Vacation" or "Civil Defense," forgetting that many students may have been too poor to have had a vacation or too engrossed in farm or school activities to have learned anything about civil defense. Finally, investigators should be mindful of a possible motivational factor in the topic assigned. How many students will write their best when asked to deal with hackneyed topics like "My Vacation" or "My Autobiography"? Some investigators have even instructed students to "Write on anything you wish. It does not matter what you write, but write until you have produced 350 words." Surely there must be some stimulating factor in a topic and, if possible, in the writing situation, too, if the writing they trigger is to have any significance for research.

---

[3] Paul Diederich wrote in 1946 that about one-fourth of a group of University of Chicago students changed their marks as a result of writing a second test essay but that less than five percent changed their marks as a result of writing a third. See his "The Measurement of Skill in Writing," *School Review*, LIV (December, 1946), 586-587. However, in a recent comment on the draft of this report, Diederich stated that two themes are "totally inadequate."

[4] Some of these considerations have been drawn from Joseph W. Miller's "An Analysis of Freshman Writing at the Beginning and End of a Year's Work in Composition" (Unpublished Ph.D. dissertation, University of Minnesota, 1958).

[5] Stephen Wiseman and Jack Wrigley, "Essay-Reliability: The Effect of Choice of Essay-Title," *Educational and Psychological Measurement*, XVIII (Spring, 1958), 129-138.

Another aspect of the assignment variable is the *mode of discourse*—narration, description, exposition, argument, or criticism. Largely ignored by people doing research in composition, variations in mode of discourse may have more effect than variations in topic on the quality of writing. Although Kincaid concluded that the writing performances of poor writers varied significantly according to the topic assigned, the fact was that his three writing assignments were very similar as topics but called for different modes of discourse.[6] His conclusion may well be reinterpreted, then, to suggest that variation of the assignment from expository to argumentative mode of discourse did not seem to affect the average quality of the writing of a group of freshmen who were better writers as much as it did a group who were worse writers. At least until such time as more research has been done on the effect of this element on writing performance, it clearly seems necessary to control mode of discourse when planning the assignments for research based on the rating of compositions.

---

[6] Kincaid, *op. cit.*

A third aspect of the assignment variable is the *time afforded for writing*. A number of studies purport to evaluate, among other things, the organization of writing when the examinees were afforded but twenty or thirty minutes to produce an essay. Although such a brief time may be sufficient for a third grader writing a short narrative on a familiar topic, it seems ridiculously brief for a high school or college student to write anything thoughtful. Even if the investigator is primarily interested in nothing but grammar and mechanics, he should afford time for the writers to plan their central ideas, organization, and supporting details; otherwise, their sentence structure and mechanics will be produced under artificial circumstances. Furthermore, the writers ordinarily should have time to edit and proofread their work after they have come to the end of their papers. It would be highly desirable to discover, through research, the optimum amounts of time needed by students at various levels of maturity to write thoughtful papers. Until such research has been conducted, investigators should consider permitting primary grade children to take as much as 20 to 30 minutes, intermediate graders as much as 35 to 50 minutes, junior high school students 50 to 70 minutes, high school students 70 to 90 minutes, and college students two hours. These somewhat arbitrary allocations of time doubtless should be adjusted according to the upper limits of the range in intellectual maturity of the students and to the topic and mode of discourse of the writing assignment.

A fourth and final aspect of the assignment variable is the *examination situation*. The situation becomes uncontrolled if the students in the experimental group all write their papers on Wednesday morning and the students in the control group write theirs right after lunch on Wednesday (when many feel logy), or the first thing on Monday (when they are still emerging from the spell of the weekend), or on Saturday morning (when they resent having to forfeit some of their weekend, even for the glory of experimentation). The time, conditions of lighting and heating, and perhaps even the popularity of the teachers proctoring the examination should be equivalent for experimental and control groups or, if improvement is being evaluated, for pretests and post-tests. Obviously the instructions given to the students should be the same, too—preferably written beforehand and read aloud to the students to prevent the inadvertent intrusion into the instructions for one group of a remark which may stimulate them more or less than the other group.

The Rater Variable

A third major variable in rating compositions is the rater variable—the tendency of a rater to vary in his own standards of evaluation. Any teacher recognizes how variable his own rating can be if he has dug some old papers out of a file, covered the grades, and regraded them without unusual care. Some of the variation may be the result of having forgotten the nature of the old assignment or the emphasis he had been making with the students back then. Although those sources of variability do not function when rating compositions for purposes of research, other familiar sources may operate and should be controlled. They may be characterized as *personal feelings* and *rater fatigue*.

Certainly the anonymity of the writer should be preserved to prevent the *personal feelings* of the rater from coloring his evaluation. That is, in a controlled experiment it should not be possible for the rater to determine from the paper in front of him whether it was written by a student in an experimental or control group. Even though the rater may not recognize the bias himself, he may be hoping that better results are obtained for one group than the other. If the rater may associate with a given group the name of the writer or of the school, the number of the class or section, or even the date on which the examination was administered, such identifying features should be removed before the papers are turned over to the rater. One way to insure anonymity is to have the students write such identifying information on a 3 x 5 card numbered with the same number as the theme paper but separated from it before the themes are submitted to the raters. Even then the numbers of the material used in the experimental groups should be so mixed with the numbers used in the control groups that the raters do not associate a continuous series of numbers with any group.

In an experiment using pretest and post-test compositions, it may be desirable not to reveal which test is which. If such an experiment is intended at all to measure improvement, concealing the identity of pretest and post-test papers is essential. Not only are the procedures mentioned above essential, but additional steps must be taken to disguise the time at which the papers were written. Students should be requested not to reveal the present year or season in what they write, and papers which do refer to "the falling leaves," "the superintendent's recent speech to the graduating class," or any other such revealing incident should be removed from the compositions to be evaluated. All the paper for both tests should be purchased and prepared at the same time to insure that differences in paper stock and printing will not be apparent. Pretests after they are written and post-tests before they are written should be wrapped lightly in brown paper and stored in the dark to prevent yellowing. The numbering of pretests and post-tests should be mixed. If the pretests become wrinkled, yellowed or musty, the post-tests should be conditioned in the same manner before being submitted to the raters. To overlook some simple identifying feature which permits the personal feelings of raters to operate may render useless all the other efforts which have gone into an experiment.

The rater variable should be controlled further by allowing for *rater fatigue*. Fatigue may lead raters to become severe, lenient, or erratic in their evaluations, or to emphasize grammatical and mechanical features but overlook the subtler aspects of reasoning and organization. Consequently, raters should not be permitted to rate late at night or for lengthy periods during the day, and they should have regular rest periods to help them maintain their efficiency. Even so, the papers should be placed in a planned sequence which does not permit more of the compositions of one group than another to be rated during a period of probable vigor or fatigue. If pretest and post-test compositions are being rated for experimental and control groups, the four types of papers must be mixed and staggered throughout the entire rating period on each day. When several readers rate the same paper (not individual dittoed or photocopied versions), no rater should place any marks on a paper; they might influence a subsequent rater. Because there are many elements which need control in the sequence of papers, it seems highly desirable to have all of the raters working in the same or adjoining offices, where the investigator can be present and, without entering into the rating himself, insure that everything runs smoothly.

The Colleague Variable

A fourth and last major variable to be considered here is the colleague variable—the tendency of several raters to vary from each other in their evaluations. The existence of this inter-rater variation has been substantiated very frequently by research. As is explained in "Objective Tests versus Actual Writing" in Chapter III, ratings of the same compositions by different raters have been found to correlate from as low as .31 to as high as .96. Consciously or unconsciously, raters tend to place different values on the various aspects of a composition. Unless they develop a common set of criteria about writing and unless they practice together applying those criteria consistently, raters may be expected to persist in obtaining low agreement.

A *common set of criteria* seems essential in coping with the colleague variable; if raters are not evaluating for the same qualities, they cannot be expected to rate with validity or reliability.[7] Three principal means of achieving this commonality are composition scales, a "general impression" method of rating, and an "analytic method."

Some forty years ago, composition scales were in wide use to standardize rating. A scale was a carefully selected set of compositions, ranging in quality from, for instance, 1 to 10. A rater would compare the paper before him to the ten sample compositions in the scale, assigning the rating of the sample composition closest in general quality to the paper in question. (The Smith study summarized in Chapter IV made use of two such scales.) The common difficulty with composition scales, however, is that the paper before the rater is seldom closely like any one of the sample compositions or that the rater notices certain similarities in which he is especially interested and overlooks or minimizes dissimilarities in other aspects of the writing. Furthermore, different scales were needed for different modes of discourse and different levels of maturity. It is easy to see why infrequent use is made of composition scales in research today. There has been a resurgence of interest in scales lately, published by universities and by state councils of English teachers, but these graded compositions seem to be designed more to help secondary school teachers develop some commonality of practice in ordinary classroom grading or to stimulate them to approximate college standards, not to help investigators rate themes for research purposes.

The two principal means of seeking valid and reliable ratings despite the colleague variable are the "general impression" method of rating compositions and the "analytic method." In the general impression method, a number of raters, working independently, quickly read and rate each composition, the mean of their ratings being used as the final rating of each paper. According to Wiseman's procedure,[8] four raters independently rate each paper, each rater "keeping to a rate of about 50 per hour" to insure that he makes up his mind quickly. Wiseman has frequently reported reliabilities in the lower .90's for raters using the general impression method for the English 11+ examinations. But the topics he reports seem to call generally for narrative writing, and the purpose of the rater is "to assess the ability of the candidate to profit by a secondary education." The general impression method may not be as effective a means of reducing the colleague variable when argumentative papers, written by older students, are being rated for research purposes.

---

[7] Stephen Wiseman disagrees with this view in "The Marking of English Composition in Grammar School Selection," *British Journal of Educational Psychology,* XIX (November, 1949), 206: "Indeed, it is arguable that, provided markers are experienced teachers, lack of high intercorrelation is desirable, since it points to a diversity of viewpoint in the judgment of complex materal, i.e., each composition is illuminated by beams from different angles, and the total mark gives a truer 'all-round' picture." But this argument seems to contain a difficulty; one would not be sure that lack of high intercorrelation was the product of diversity of viewpoint or the product of erratic marking.

[8] *Ibid.,* p. 208.

In the analytic method, two or three raters independently assign a number of points to each of several aspects of a composition and total the points to obtain an overall rating, which is then averaged in with the overall ratings of the other raters. More time-consuming than the general impression method and hence more expensive if two or more raters are used, the analytic method does have the advantage of making clear the criteria by which the rating is done.

In a comprehensive research into four different methods of rating compositions, Cast found the general impression and analytic methods more reliable than the other two and the analytic method slightly superior to the general impression method.[9] Acknowledging that, when used by a trained and experienced rater, the general impression method may correct the errors to which "a crude, mechanical, quantitative dissection might inevitably lead," she concluded that the analytic method, "though laborious and unpopular, appears almost uniformly the best" and that the unreliability of rating "can evidently be greatly reduced by standardized instructions and by the training of examiners."

A caution must be made about the analytic method, however. The criteria used in an analytic method must be clearly defined. In one scheme, the general effect is that half of the total rating is ill-defined:

Quantity, Quality, and Control of Ideas	50	marks
Vocabulary	15	”
Grammar and Punctuation	15	”
Structure of Sentences	10	”
Spelling	5	”
Handwriting	5	”
Total	100	marks[10]

---

[9] B. M. D. Cast, "The Efficiency of Different Methods of Marking English Composition," *British Journal of Educational Psychology*, IX (November, 1939), 257-269, and X (February, 1940), 49-60.

[10] P. Hartog and E. C. Rhodes, *The Marks of Examiners* (London: Macmillan Company, 1936), p. 138.

To turn that analytic scheme into a meaningful system, one would have to divide or define in more detail the first category in the list. Although different in emphasis because designed for the writing of college freshmen, the theme examination criteria used at the University of Iowa seem to offer a better balance of considerations, especially when they are seen in the light of the three-page set of instructions defining each category:

Central Idea and Analysis	1-5	points
Supporting Material	1-5	"
Organization	1-5	"
Expression (diction and sentence style)	1-5	"
Literacy (grammar and mechanics)	1-3	"
Total Possible	5-23	points[11]

There is a danger in any analytic system that a beginning rater will first establish the total number of points according to his general impression of a composition's merit and *then* apportion the total points among the various categories so that they add up to the total. Such a practice, of course, undermines the basis of the analytic method and shows the need for what Cast called "the training of examiners."

Some substantiation of the importance of *practice rating* was provided by Stalnaker, who had an undisclosed number of college English instructors carefully reread a composition examination after a period of training. He found that rater reliability on the first reading was as low as .30 and never as high as .75 but that, after training, the reliabilities on the second reading ranged from a low of .73 to a high of .98 with an average of .88.[12] Although the unusual nature of the examination (it included the construction of an outline and the revision of sentences, among other things) prevents Stalnaker's study from constituting conclusive proof of the efficacy of rater training for the grading of compositions, his findings are reinforced by the frequency with which rater training is reported in studies achieving high reliabilities. A caution must be offered, however. Even though raters are requested to consider in their evaluations such attributes as content and organization, they may permit their impressions of the grammar and mechanics of the compositions to create a halo effect which suffuses their general ratings. (A converse emphasis, of course, can just as easily create the halo.) Evidence of such a grammar halo effect has been offered in at least two studies, one by Starring[13] and the other by Diederich, French, and Carlton.[14] It must be noted that Starring's raters (in contrast to Diederich's) used an analytic method and had had regular practice theme rating sessions, though it was his impression that the sessions had not produced much agreement. Perhaps one way that the rater variable can be further controlled is to use the ratings on common practice themes as a basis for pairing raters with differing standards of severity and leniency. But the effectiveness of this practice evidently has not been investigated in research.

---

[11] The 5 represents "A," 4 "B," and so on to 1 "F." If a student receives an "F" in any one of the five categories, his paper fails.

[12] John M. Stalnaker, "The Construction and Results of a Twelve-Hour Test in English Composition," *School and Society*, XXXIX (February 17, 1934), 218-224.

Probably the basis for effective use of the common set of criteria in an analytic system lies in the commitment which each rater feels toward the criteria being employed. If he has shared in developing the criteria or had an honest opportunity to share in revising them (as the graders did in Buxton's study,[15] reported in Chapter IV), he ordinarily should be expected to enter into practice rating and actual rating with an honest effort to make the method work. Even so, periodically during the actual rating (Buxton's graders did it with every twenty-fifth paper), the graders should jointly review a composition they have just rated, insuring that they are maintaining a common interpretation and application of the criteria they are using.

If this analysis of the four major variables in rating compositions is discouraging and if the procedures for controlling the variables seem complex, it is because composition itself is complex and the rating of it a challenge. But, if English teachers are to do more than "speculate about the conditions for teaching composition," investigators must plan and carry out the rating of compositions so that the major elements are controlled. To do less is to waste one's research efforts. There is an alternative, however—the alternative procedure of analysis employed in the Harris study (reported in Chapter IV).[16] One may use frequency counts.

Frequency Counts

Typically, frequency counts have been a most frustrating type of composition research to read, yet perhaps they are to become one of the most important types, as exemplified by the Harris study. The importance of the frequency count (in contrast to rating procedures) lies in its potential for describing a composition in fairly objective terms which *can* mean the same things to most teachers and investigators and which are subject to more statistical analyses than are ratings. The frustration comes from confusion over the purpose of such studies and from failure to use methods meaningful to other investigators. A review of some of the methods used may clarify the point. Suggestions for improving the value of such studies are placed in italics.

---

[13] Robert W. Starring, "A Study of Ratings of Comprehensive Examination Themes When Certain Elements Are Weakened" (Unpublished Ed.D. dissertation, Michigan State College, 1952).

[14] Paul B. Diederich, John W. French, and Sydell T. Carlton, *Factors in Judgments of Writing Ability*, Research Bulletin RB-61-15 (Princeton: Educational Testing Service, 1961).

[15] Earl W. Buxton, "An Experiment to Test the Effects of Writing Frequency and Guided Practice upon Students' Skill in Written Expression" (Unpublished Ph.D. dissertation, Stanford University, 1958).

[16] Roland J. Harris, "An Experimental Inquiry into the Functions and Value of Formal Grammar in the Teaching of English, with Special Reference to the Teaching of Correct Written English to Children Aged Twelve to Fourteen" (Unpublished Ph.D. dissertation, University of London, 1962).

Many investigators have counted and reported the total numbers of errors of various types which they have found in a collection of compositions. Usually, the errors they have sought have been errors in grammar, usage, and mechanics. If an investigator is seeking examples of pronoun disagreement, for instance, he makes a tally on a sheet every time he sees an infraction of the rule he has in mind. One difficulty with many such error counts is that the reader does not know what "rule" the investigator has in mind. Is he counting as an error "Everybody went back to the classroom and got their books"? Or does he accept that construction as a nonerror? Does he count "It's me" as a nonerror, an error in pronoun agreement, a problem in the predicate nominative, a failure in case agreement, or simply an example of "poor diction" or even "unidiomatic usage"?[17] *It is essential for the investigator to give clarifying examples for each type of item he is counting.* But even then the reader may feel some hesitation about the results; it is very difficult in a few examples to reveal clearly the many decisions which must be made in classifying instances of disputed and changing usage.

The more thorough the investigator, the more he may subdivide types of errors into lesser categories. Some error counts distinguish among more than 400 types in this fashion, while others may divide the same problems into but 30 types. Such variation makes it impossible to compare one study to another or to synthesize their results. *If frequency count studies are to be useful to other investigators, then, they should be based on a standard classification of types of items.* There is no generally accepted standard classification at this time.

Thirty years ago, one writer constructed a composite list of "the most common grammatical errors," drawing from 33 previous error counts. The absurdity of the list is apparent today not only because the categories of the 33 studies had been different but because the counts were made from compositions on various topics, in differing modes of discourse, and by children and adults of widely varying maturity and ability who came from various dialect regions and socioeconomic backgrounds. The composite list even lumped together error counts of oral and written language. It is appropriate to ask what the purpose of such a list would be. If it were to help English teachers and curriculum-makers determine which features of grammar need to be taught, then the frequency count should be conducted from the writing of the pupils to be taught, or from pupils similar to them at the same grade levels. If it were to help determine which types of grammatical items should go into a college English placement test, then probably the frequency count should be based on the writing of a cross section of freshmen or upperclassmen at the types of colleges in which the test will be used. If the count were to be used to establish national norms in the development of written exposition from grade 1 through grade 12, then the compositions would have to be selected from among expository papers written by "average writers" at each grade level and sampled from groups of various socioeconomic backgrounds, amounts of writing practice, and geographical areas.

---

[17] One investigator conducted two error counts of the same paragraphs, employing a conservative approach to usage on one occasion and a liberal approach on the other. Although the two counts yielded the same results for such matters as spelling and capitalization, the two counts differed markedly for such matters as misuse of pronouns. See page 73 in Hugh N. Matheson's "A Study of Errors Made in Paragraphs . . ." (Unpublished M. A. thesis, University of British Columbia, 1960).

The studies by Kincaid (summarized in Chapter IV) and by Wiseman and Wrigley[18] demonstrated that the topic a person writes on affects the caliber of his writing. Seegers has indicated that a person's sentence structure is affected by the mode of discourse he is using—argumentation, exposition, narration or description.[19] It also seems apparent that a pupil's rhetoric, syntax, and usage vary to some degree with his general ability, experience in writing, maturity, socioeconomic background, and native geographical area. Consequently, *before conducting a frequency count or using the results of one, a person should determine what his purpose is and then ascertain that the compositions used are appropriately controlled or sampled according to topic, mode of discourse, and characteristics of the writers.*

A fundamental difficulty with most frequency counts is that they are simply counts of grammatical and mechanical "errors," omitting attention to purpose and main idea, supporting material, organization, and style. Even though, in the "summary, conclusions, and implications" chapter of his thesis, the investigator expresses regret at the impossibility of counting rhetorical elements, the impact is often unfortunate; the study has distracted the investigator, his major professor, and readers of the report from the "larger elements of composition." It is obvious that soundly based counts are needed of the frequency of various grammatical, word, and mechanical usages; but even more urgently needed are *similar analyses of rhetorical constructions.*

---

[18] Wiseman and Wrigley, *op. cit.*

[19] J. C. Seegers, "Form of Discourse and Sentence Structure," *Elementary English Review,* X (March, 1933), 51-54.

Imaginative approaches to frequency counts are needed. The tendency in any frequency count is to find what one is looking for. More investigators need to initiate frequency studies with fresh questions in mind, not merely attempting to find new frequencies of old "errors." Some psychologists have been trying new approaches. Kimoto, for instance, explored the relations between dominance-submissiveness characteristics and grammatical constructions.[20] She asked a number of subjects how they would respond in each of several situations in which their own tendencies to dominate or submit would be tested. After recording their oral responses, she counted the frequency of such grammatical features as the passive voice and discovered a number of interesting things. Although her study is not very germane here, it does exemplify an approach which may open up new dimensions in the teaching and learning of composition. Investigations have also been made, using frequency counts, into the degree of abstractness of writing,[21] the correlates of egocentricity,[22] some variations in style, [23] abstraction as an index to linguistic maturity,[24] and the increased use of subordination with maturation.[25] These studies have all tended to be exploratory in nature, attempting to develop new instruments for the analysis of language. The worth of such instruments becomes better known, of course, when other investigators attempt to validate the instruments. For instance, Haskins validated the Gillie abstraction index by measuring the degree of abstraction of the articles in an issue of the *Saturday Evening Post* and then comparing the reactions of a "nationwide sample of readers (N = 340)" to the abstraction of the articles.[26] Although he did not explain the basis for selecting his sample of readers and he accepted their simple statements about which articles they had read and found satisfaction from, Haskins' article does give one more confidence in the Gillie formula. A study by Anderson attempted to validate several frequency count instruments.[27] Although Anderson points out that his own use of 150-word samples of writing was a weakness in his study, he does show, for instance, that the widely known LaBrant subordination index does not work well if not applied under carefully prescribed conditions.

---

[20] Blanche Kimoto, "A Quantitative Analysis of Grammatical Categories as a Measure of Dominance" (Unpublished Ph.D. dissertation, Boston University, 1951).

[21] Paul J. Gillie, "A Simplified Formula for Measuring Abstraction in Writing," *Journal of Applied Psychology*, XLI (August, 1957), 214-217.

[22] John A. Van Bruggen, "Factors Affecting Regularity of the Flow of Words During Written Composition," *Journal of Experimental Education*, XV (December, 1946), 133-155.

[23] David P. Boder, "The Adjective-Verb Quotient: A Contribution to the Psychology of Language," *Psychological Record*, III (March, 1940), 310-343.

[24] Gustav Kaldegg, "Substance Symbolism: A Study in Language Psychology," *Journal of Experimental Education*, XVIII (June, 1950), 331-342.

[25] Lou L. LaBrant, "A Study of Certain Language Developments of Children in Grades Four to Twelve, Inclusive," *Genetic Psychology Monographs*, XIV (November, 1933), 387-491.

[26] Jack B. Haskins, "Validation of the Abstraction Index as a Tool for Content-Effects Analysis and Content Analysis," *Journal of Applied Psychology*, XLIV (April, 1960), 102-106.

One way to break from the grip of error counting is to *count the frequency of certain types of situations and the ways in which writers of various kinds respond to those situations*. For instance, instead of merely counting what he happens to consider errors in the "these kind of things," "these kinds of things," "this kind of thing" expression, the investigator would do well to tabulate the frequency of each of the ways in which writers meet this situation (as Thorndike did[28]) and to seek correlations of the type of response and the type of writer (age, amount of experience in writing, general writing ability, socioeconomic background, and geographical area). Not only would such data help determine what usage label could be attached to each type of response, but, unlike counts of errors, the data would be meaningful even when usage is disputed or when notions of "correctness" have changed since the study was conducted. Such descriptions of actual usage would be more soundly based than the questionnaire approach employed by Leonard and many others who merely asked people which of several expressions they used.[29]

The reporting of frequency counts has often been meaningless or confusing because of the way in which the data have been expressed. The earliest counts seem merely to have reported the *total number of errors* found in the writing examined. All things being equal, if a person tabulated apostrophe situations in 200,000 words, he would find twice as many situations as if he had examined 100,000 words. To overcome this difficulty, some investigators reported their results by listing the *errors in rank order of frequency*. But this procedure had two shortcomings. It hid the actual frequency behind the rank; a reader could not tell whether an error of the first rank was much more prevalent or barely more prevalent than an error of the second rank, etc. It also hid the actual frequency in cases where many errors increased or decreased from grade to grade or before and after an experiment even though the relative frequency, or rank, remained fairly constant.

---

[27] John E. Anderson, "An Evaluation of Various Indices of Linguistic Development," *Child Development*, VIII (March, 1937), 62-68.

[28] Edward L. Thorndike, "An Inventory of English Constructions with Measures of Their Importance," *Teachers College Record*, XXVIII (February, 1927), 580-610.

[29] Sterling A. Leonard, *Current English Usage* (Chicago: Inland Press, 1932).

A third means of report was the *error quotient;* that is, the number of errors of a given type divided by the number of opportunities to make that type of error. Like the rank order, the error quotient, reported without supporting data, hid the actual frequency. If a composition contained one semicolon error in two opportunities to err, and five comma errors in ten opportunities for them, the error quotient was the same; the higher frequency of comma problems was not apparent. Similarly, the actual frequency was hidden by the *percentage of errors*—a percentage computed by dividing the number of errors of a given type by the number of errors of all types. Moreover, since the number of errors of all types varies considerably according to the number of types of errors being considered, the percentage of errors is especially meaningless.

It is obvious that frequency counts cannot be very meaningful if their results are expressed in relation to an indefinite concept like total number of errors. A definite, indisputable number is needed, one which clearly relates to the actual frequency with which an item occurs in writing. *The frequency of a type of item should be reported per hundred or thousand words.* That is, the total actual count of a particular type of item should be divided by the total number of hundreds or thousands of running words of composition from which the count was tabulated. The only condition necessary is that a standard system be employed for counting the total number of words, such as that used by Chotlos.[30]

One further precaution needs mention. If one bases a frequency count on ten samples of 50 words each, a total of 500 running words, he doubtless does not have a sample representative of anything. If he bases his count on 1,000,000 running words, however, he may be counting much more than he needs to in order to arrive at a stable frequency. The number of running words needed depends on the frequency with which the situation occurs and the care with which the investigator controls or samples various topics, modes of discourse, and characteristics of writers. The most careful way to determine the number is to *use the "cumulative-average" technique of sampling*[31]—drawing successive samples of the same number of running words until the cumulative averages reach a relatively stable point for most of the frequency situations being considered.

---

[30] John W. Chotlos, "A Statistical and Comparative Analysis of Individual Written Language Samples" (Ph.D. dissertation, University of Iowa, 1942), pp. 14-19. Published in *Psychological Monographs,* LVI (1944), 77-110.

[31] James W. Evans, "The Social Importance and the Pupil Control of Certain Punctuation Variants" (Unpublished Ph.D. dissertation, University of Iowa, 1939).

The suggestions made thus far for the improvement of frequency counts all seem to add up to one thing—more tedious work for the investigator. This is not necessary. Instead of conducting fishing expeditions in a morass of 400 types of items, *more investigators employing frequency counts should focus their studies on narrower, more clearly defined areas and explore them more thoroughly and carefully.* The same amount of effort should be employed in more intensive analyses of more limited problems.

Another means of efficiency should be explored. Instead of counting many different types of items to study a larger area of concern, *investigators should seek to discover certain key situations which are indices of larger areas of concern.* For example, would not the usage of irregular English verbs provide an index to the general level of usage of primary school children? Does the subordination index truly provide an index to a broader aspect of linguistic development in writing, as the Strickland study shows in speech?[32] Over the years and through the cumulative efforts of many investigators, if a number of key indices can be developed, frequency counts may become a very efficient means of studying written composition.

General Considerations

The consideration of "Rating Compositions" and "Frequency Counts" dealt in detail with two concerns unique to research in written composition. Here some more general suggestions will be offered on designing and reporting research in composition, drawn up as the writers of this report, especially the specialist in educational research, noted the strengths and weaknesses of the studies being reviewed. Helpful in writing this section were articles by Carroll,[33] Dolch,[34] Rivlin,[35] and Singleton.[36]

The following discussion is not addressed to investigators as a substitute for formal study of research design and statistics; rather the discussion is intended to introduce the reader of research to some basic considerations to have in mind as he interprets and evaluates reports.

---

[32] Ruth G. Strickland, "The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children," *Bulletin of the School of Education*, Indiana University, XXXVIII, 4 (July, 1962), 1-131.

[33] John B. Carroll, "Neglected Areas in Educational Research," paper presented at a meeting of the American Educational Research Association, Chicago, February 23, 1961. Mimeographed.

[34] E. W. Dolch, "School Research in Reading," *Elementary English*, XXXIII (February, 1956), 76-80.

[35] Harry N. Rivlin, "The Present Status of Research in Functional Grammar," *English Journal*, XXVII (September, 1938), 590-597.

[36] Carlton M. Singleton, "Freedom to Research," *Elementary English*, XXXVIII (February, 1961), 114-117, 121.

Attitude of the Investigator

One day a doctoral candidate consulted the director of a large freshman English program. The graduate student had heard that the professor had "a lot of data" on the two thousand freshmen who went through the program each year, and he wondered if the professor would permit him to use it. "I'd like to do something on reading," the student explained, "–on the effects of reading on composition." He elaborated: "If I could get the grades the students receive in your courses and compare them to the amount of reading they report having done in high school, I'd be able to tell how much effect high school reading has on composition and be able to recommend that the high schools stimulate more reading."

The student had made a number of assumptions which needed–and received–questioning. Here are some of them:

1. That course grades in freshman English reveal the quality of a student's writing. (But a student's course grade may have been lowered for failure to do assignments, for cutting class, for poor work in the reading aspect of the course, for failure to take part in enough class discussion, for being personally obnoxious, for being in the class of a severe grader, etc. The grade could have been raised for doing unassigned work, for having other students write his papers for him, or for the converse of the other matters mentioned above.)
2. That these college freshmen are generally typical of high school students. (But many high school students have ability, finances, or motivation too low to bring them to college. Furthermore, those high school graduates attending this particular university doubtless are not representative of all high school graduates who are college freshmen.)
3. That the freshmen can remember and would report accurately the amount of reading they had done in high school. (But many freshmen might, consciously or unconsciously, report more reading than they actually did.)
4. That "amount of reading" is a clear concept. (But it is not clear whether or not "amount of reading" refers to number of books regardless of size, number of pages, books well understood or

books vaguely understood, magazines and newspapers as well as books, books checked out of the high school library whether or not read, etc.)

5. That "amount of reading" is equivalent to "reading." (But the nature of the things read, of the attitude the students held toward their reading, or of their comprehension of the reading may be more important than the amount of reading they did.)
6. That a correlation between amount of reading and quality of writing would reveal a causal relationship. (But these two matters could be caused by some third, unknown factor, or several factors, such as the intellectual atmosphere of the students' homes or school requirements that certain amounts of reading and writing be done.
7. That he knew what the results would be before he had made the investigation. (But such a bias could color his entire investigation, interpretation of results, and recommendations based on the results.)

It is clear that research must be carefully designed if it is to be effective. Basic to a good design is the honest desire to discover or test some generalization about which the investigator does not believe he is fully informed, to discover or test some answer to a sincere question. Coupled with that honest search for knowledge should be a rather antithetical unwillingness to believe anything without being shown, moderated a little by the realization that some things cannot be shown as conclusively as others.

Meaning of Terms and Measures

Basic to honesty is clarity. Terms and criteria may mean nothing in the abstract. It should be clear what they represent. If a composition is being rated in part for "fluency," for example, the meaning of that term should be made clear. It could refer to the number of words a student writes, the speed with which he writes, writing without correcting or adding elements, or even writing so that the reader proceeds smoothly from one idea to the next. Terms and criteria should be defined carefully, preferably in an operational manner, permitting others to use the terms and criteria with the same results. It is not enough, for example, to refer to the amount of predication a student uses if it is not clear whether "glanced covertly and winked suggestively at him" represents one predicate or two.

Behavior should be studied by direct observation and measurement, not only by indirect methods. It is not enough to ask a student to identify a logical fallacy in the writing of others if the problem is whether or not the student uses good logic in his own writing; he should be given an assignment which directs him into writing where his own valid or fallacious logic will be displayed. Similarly, it is not enough to ask a student whether or not he constructs a written outline before he writes; it would be much preferable to give him scratch paper with his theme paper and examine both afterward to see whether the student has written any notes which can be termed an outline. Investigation should not depend upon such indirect measures as multiple-choice or true-false tests, questionnaires, self-inventories, and the like.

Statistical analyses in composition research are based upon criterion measures about which certain assumptions must be made. The nature of these assumptions should be made clear, and there should be fairly adequate evidence that the assumptions are valid and that the criterion measures can be applied reliably. For instance, if an investigator wishes to measure originality of diction by determining how many words in a composition are not on a list of 10,000 most frequently used words, he must be prepared to demonstrate that "infrequently used words" usually yield diction which recognized authorities would characterize as "original." Furthermore, he must define what is meant by "not on the list" so that other investigators can analyze the same compositions with the same results; he must explain, for instance, how to count a word which has a different suffix from a word on the list.

Planning of Procedures

As the anecdote about the graduate student showed, good research generally does not consist of an after-the-fact description or mere correlation study of the results of behavior over which the investigator had no real control. (Correlation studies are necessary, of course, but the investigator should be wary of accepting a correlation as a cause.) He should plan his groups and treatments before he gathers his data, specifying before beginning his study what statistical procedures he will use. Then he should be reasonably consistent in his use of those procedures. He should not, for example, switch criterion measures in different parts of an experiment when there seems to be little basis for the change, as did the investigator who used the subordination index as a measure of language development from grade four to grade eight but then depended upon sentence length as his measure through grades nine to twelve. The statistics need not be complicated, but they should be appropriate. For instance, the progress of a range of students should not be examined only by mean scores, when average gains may be achieved merely by speeding up one end of the range; distributions of scores should be examined.

Controlling of Variables

If the investigation entails the comparison of one method of instruction to another, all variables other than the method should be controlled —the personality, knowledge, and experience of the teacher; the mental ages, writing proficiency, and socioeconomic and intellectual home backgrounds of the students. The attitudes of teachers and students should be controlled (that is, selected) before the experiment and measured after it.[37] It is frequently better to control such variables than to choose teachers and students at random and hope for the best. If students at the extremes of a range of proficiency on one measure are used, it is wise to allow for a regression effect in other areas; that is, just because a group of generally good writers is being contrasted to a group of generally bad writers does not mean that the two groups are far apart in spelling ability or in attitudes toward reading. In fact, the likelihood is that they will be closer on all other measures than they are on the one used to differentiate them. It is better to select subjects than to use volunteers, because only certain kinds of people are willing to volunteer. Moreover, the students should be chosen in such a way that they represent some meaningfully defined student population; otherwise, the results of the experiment cannot validly be generalized beyond those involved in the experiment. Enough students and teachers should be involved that unobserved, uncontrolled factors may cancel out each other.

Other "outside influences" should be controlled or otherwise accounted for—the time of day classes meet, motivation by classroom guests or rewards, size of classes, and demands upon time and initiative. A startling example of outside influences was noted in an experiment in which 24 third and fourth graders were being taught to typewrite. The effect of this instruction on written composition was being compared to the effect of no such instruction on the written composition of a control group. The experimental boys and girls came to a university campus during the summer, were instructed in a room newly redecorated for the purpose and furnished with new desks adjusted to each child's height. A new portable electric typewriter was provided for the use of each child, the typewriters of different pastel colors being switched from child to child each two weeks. Four assistants were in the room to help the professor who taught the class, among other things demonstrating typewriting finger and wrist movement at the piano keyboard. Newspaper publicity covered the experiment throughout that area of the state. Meanwhile, what was happening to the 24 third and fourth graders in the control group? The published report did not say. This experiment illustrates not only the influence of "outside" factors but of the "Hawthorne Effect," the added stimulation received by an experimental group when a new method is being compared to an old method (or, more likely in this case, no method).

---

[37] A constructive and understanding attitude toward an experiment may be generated and pitfalls may be avoided if the investigator explains the experiment beforehand to the teachers and students who will participate in it, soliciting their suggestions and utilizing them when possible. Of course the same treatment here must be afforded experimental and control groups.

Some additional influences should be watched for. If students are doing two or more things in a sequence, the sequence effects must be controlled. For example, if oral and written samples of the same narrative are being collected, some students should first speak and then write, and other students should first write and then speak. If a procedure or instrument is being used which would not be employed in a regular teaching situation (such as a kymograph, recording on a moving drum the starts and stops of a student's writing), steps should be taken to insure that the atypical element did not affect the outcome of the experiment. And, finally, the influence of time and disuse should not be ignored, as it usually is in composition research. Often a follow-up measure should be taken, months or even a year after a new method has been tried, to see how learning stands up, for experimental and control groups, when instruction and practice lie in the past. And then the investigator should take steps to determine if one group has advanced through further study or practice, while the other group retrogressed through disuse, and to determine whether or not these differing post-experiment behaviors were generated by attitudes developed during the experiment.

Need for Trials and Checks

This cursory review of elements in the design of composition research should make one point clear: there are many elements to measure and control, and unexpected influences which can spoil a carefully designed experiment. For that reason, it is very prudent to have a trial run before actually beginning an experiment. Different teachers and students must be used for the trial, of course, to prevent the trial itself from affecting the subjects in the main experiment. The complexity of composition research also makes it prudent for the investigator to check on the progress of the experiment while not himself becoming an outside influence. Rather than assume, for example, that the teachers of his experimental and control groups are following the agreed instructional procedures, he should periodically check, by observation or conference, to see that things are proceeding as planned.

Reporting of Results

Just as one major flaw in design can render an otherwise careful investigation inconclusive, so can a major omission from the report make an investigation *seem* worthless. The writers of this pamphlet were forced to set aside a number of studies which seemed to have employed sound statistical procedures but which did not make clear what the statistics represented. The most frequent instance is the comparison of two methods which affords no descriptive details of the methods, merely a few generalizations which have different meanings for different people. It is essential in such cases that instructional procedures be described clearly (or made implicit in textbooks, exercises, films, machines, etc.) so that other investigators may reproduce at least the general nature of the experiment or may replicate it in its entirety. Similarly, the nature of the students must be described in enough detail to permit the reader to determine for which kinds of students the results are applicable, and the investigator must be careful not to generalize his conclusions beyond the limitations of the type of population he sampled from. The meaning of the raw data should be made clear by the inclusion of sample themes (with markings), tabulations, distributions of scores, and the like. Such raw data should often be included for individuals and/or defined groups of individuals, not merely for the group as a whole, permitting the reader to determine whether or not average gains were achieved merely by speeding up one extreme of a distribution of students. Although averages and percentages are necessary to make data meaningful, such measures and elaborate statistical methods should not obscure the raw material of the investigation. Terms like "high" and "low" should be avoided in referring to reliability coefficients, because such labels are very subjective. Nor should nonstandard procedures of quantitative and statistical description be used. In short, the data should be described and analyzed by methods which permit a clear understanding by the reader and replication by other investigators. And the lay reader should be reminded that statistically significant results may have very limited practical significance.

Finally, if the report is a thesis or dissertation which will be available on microfilm, the investigator should make certain allowances for the medium used. He can assist the reader by including an abstract at the beginning of the manuscript and making the table of contents detailed enough to permit ready location of subsections of chapters. Furthermore, he should avoid the unnecessary use of pages too wide to read in the microfilm reader and the use of color in graphs which become meaningless in the black and white medium.

III.

THE STATE OF KNOWLEDGE ABOUT COMPOSITION

Some months ago, one of the writers of this report mentioned to a colleague doing research in internal medicine that it was disappointing to see how little was really known about the teaching and learning of written composition, how inconclusive most of the research has been. The colleague replied that 95 percent of the research in his area was inconclusive or trivial. "Keep at it," he said. "As you learn more, you'll slowly learn to define your problems in a useful manner and to refine your techniques of analysis. Then you'll be in a position to learn something substantial." His emphasis on the importance of research goals and procedures is reflected in this report; much of what is known about composition teaching is actually known about the procedures of research—and has been considered in Chapter II.

The purpose of Chapter III, however, is to review what is known about the teaching and learning of written composition, as distinct from research procedures. Although the emphasis of this chapter is on the highly selected research summarized in Chapter IV, reference is also made to numerous other studies which merit attention. The findings of these secondary studies should not be taken as conclusive, however. Sometimes the studies make claims which go far beyond the statements extracted here. Other times the investigators base their conclusions entirely on objective testing rather than on actual writing, or they themselves do the teaching in their experiments or rate the themes on which their results are based. Such cautions are not intended to disparage the secondary studies so much as to remind the reader that there is more to be discovered on the subject, as the authors of these studies often state in their own reports.

Environmental Factors Influencing Composition

The primacy of the writer's own experiences has been explored in two interesting, similar studies—by Anderson[1] and Edmund[2]—of the relationship between prior experiences and short story writing. In each case the investigator discovered, by means of a questionnaire, which short stories (or the settings and characters employed, in the Anderson study) were based on derived experience. Anderson's study used 42 short stories written out of class by freshmen—many of them war veterans—in the investigator's college composition class, and Edmund's study was drawn from the in-class writing of 90 seventh graders in three suburban New York State schools. Although both investigators were involved in evaluating the stories (Anderson did all the evaluating; Edmund evidently was one of the three judges) and certain other problems cast doubt on the significance of the results, still the general problem and some of the procedures of the studies merit further investigation. With these cautions, it is interesting to note that Anderson's college freshman veterans seemed to write better narration when they based it on direct experiences, Edmund's seventh graders when they drew their ideas from such derived experiences as television programs and motion pictures.

---

[1] Edward L. Anderson, "A Study of Short Stories Written by Students in College Composition Classes to Determine Relationships Between the Prior Experiences of the Students and Their Treatment of Setting and Character" (Unpublished Ph.D. dissertation, New York University, 1950).

[2] Neal R. Edmund, "A Study of the Relationship Between Prior Experiences and the Quality of Creative Writing by Seventh Grade Pupils" (Unpublished Ed.D. dissertation, Syracuse University, 1956).

McClellan also conducted a study of "creative writing," but with a different purpose and at a different level.[3] His data were 200 papers randomly selected from those of 1,065 children in grades 3 through 6. Of particular interest is the investigator's study of the socioeconomic level of the children. They were attending three schools representative of the upper, middle, and lower economic classes, and he found that, with almost every factor studied, the higher the socioeconomic level the better the performance.

An indirect key to environmental factors influencing composition is the composition interests of school pupils. If we know what they want to write about, we may be able to approach them on their own ground more readily. There is much research on the reading interests of school pupils, but little has been done on their writing interests. Although now too old to provide reliable clues to writing interests, Coleman's study affords some leads—and cautions—to those who may wish to attempt an investigation of current interests.[4] The weaknesses of Coleman's study—and his may be the best of those few done—reflect the difficulty of this type of research: (1) The teachers were not requested scrupulously to avoid influencing their pupils' choices of topics for writing. (2) The papers were classified as to interest evidently by title only, not by the actual content of the papers. (3) Most basic of all, distinct and mutually exclusive classifications were not developed, suggesting that it was often arbitrary in just which classification of interest a composition was placed. For instance, "war" was not one of the categories used. Of those used, under which would a war story be placed—"Adventure," "Contemporaneous Famous People," "Historic Events, Sites, or Characters," "Outdoor Activities," "People," "Sympathy," or "Social Problems"? (4) The pupils' preferences for forms of discourse were sought by questionnaire, not by examination of the papers they wrote, and the nature of some items on the questionnaires may have tended to encourage some types of responses and discourage others. (5) The investigator did not attempt in any systematic way to discover the intensity of the pupils' interests or the relation of their interests to such factors as IQ and socioeconomic level. In short, the area of pupils' writing interests is wide open for more research, but the problems involved are prodigious. Probably an entirely fresh approach is needed if progress is to be made in this area—an approach which goes deeper than the superficial concerns of the student.

---

[3] Jack McClellan, "Creative Writing Characteristics of Children" (Unpublished Ed.D. dissertation, University of Southern California, 1956).

[4] J. H. Coleman, *Written Composition Interests of Junior and Senior High School Pupils*, Contributions to Education, No. 494 (New York: Bureau of Publications, Teachers College, Columbia University, 1931).

One study which seems to probe into the psychological realm underlying or accompanying the act of composition is Van Bruggen's investigation into the flow of words.[5] In an experiment using a kymograph, Van Bruggen measured the "rate of flow" of words while junior high school pupils were writing compositions. He sought "to determine how it [rate of flow] is affected by various compositional, academic, personal, and environmental factors" and "to discover how the composing structure—that is, the number, length, and location of pauses between words—differs in compositions of superior and inferior quality and in compositions written with rapid and slow flow of words." Although the investigator took careful precautions to see that use of the kymograph did not obtrude on the students' writing, some of his procedures for analyzing the compositions seem very questionable, and some of his conclusions seem to leap beyond the reasonable distance for a step from data to generalization. But this study of the number and length of pauses between words may provide leads and procedures for valuable exploration of the psychological dimension of the composition act.

The Van Bruggen study, among others, suggests to the writers of this report that the psychological dimension of writing needs to be investigated by case study procedures. Individual differences may

---

[5] John A. Van Bruggen, "Factors Affecting Regularity of the Flow of Words During Written Composition," *Journal of Experimental Education*, XV (December, 1946), 133-155.

"cancel out" in studies using the mean as the measure of a group. Case studies have done much to help remedial reading specialists understand and assist their "clients," and the similar complexities of writing suggest that much may be gained by developing case study procedures, against a background of experimental group research, to investigate the factors affecting the learning of composition and the procedures which will accelerate and maintain learning. But before composition teachers can conduct case study investigations, they must learn how to do so.

Another promising type of investigation is the longitudinal study—the type of study which follows the same individuals through a protracted period of time. The longitudinal study is especially appropriate for written composition, in which change usually seems to take place slowly. Note the effectiveness of the Harris study (summarized in Chapter IV) because it extended for two years instead of one.[6] But the Harris study would not ordinarily be termed a longitudinal study. A clear example is the Loban project, which began in September, 1952, to study the development of language ability of 338 children in eleven kindergarten classes from socioeconomically diverse districts in Oakland, California. Working with a team of assistants and financed by a grant from the U. S. Office of Education, Loban has analyzed oral language samples obtained yearly from his subjects. He has also obtained written samples, beginning with the third grade (a written response to a colored picture), but, so far at least, has subjected the writing only to rough evaluation for the light it may throw on the development of oral language. Here is one finding from his analysis of oral language: "Those subjects who proved to have the greatest power over language . . . were *the subjects who most frequently used language to express tentativeness.* Supposition, hypothesis, and conditional statements occur much less frequently in the language of those subjects lacking skill in language."[7] Numerous investigators have discovered that the incidence of adverbial clauses of concession, though small, becomes more apparent in a somewhat consistent pattern as one analyzes the writing of older children. If one were to teach students, over a period of years, to express "tentativeness" (among other things) in their writing, would he help them develop their writing ability more than they otherwise would? Such questions must be explored in the context of a longitudinal study, not in a nine-month experiment.

---

[6] Roland J. Harris, "An Experimental Inquiry into the Functions and Value of Formal Grammar in the Teaching of English, with Special Reference to the Teaching of Correct Written English to Children Aged Twelve to Fourteen" (Unpublished Ph.D. dissertation, University of London, 1962).

[7] Walter Loban, *Language Ability in the Middle Grades of the Elementary School* (Final Report to the U. S. Office of Education, Department of Health, Education, and Welfare, 1961), p. 89. Mimeographed. Also reported in Walter Loban, *The Language of Elementary School Children* (Champaign, Illinois; National Council of Teachers of English, 1963).

Obviously a question like that above is most profitably investigated over a protracted period of time by or under the direction of an experienced research person who understands students and the written language, not by a loosely guided Ph.D. candidate hurriedly trying to complete his degree requirements soon so that he can adequately feed, clothe, and house his family. There are other problems involved in conducting longitudinal studies. The principal investigator must be mature, but he cannot be old if his project is to extend for twelve years. He must be able to develop measures and procedures which will be effective at all the levels of maturity through which his subjects will pass. He must be able to elicit regular financial support from foundation, governmental agency, or university. He must be able to direct the others who will work with him and obtain the continued cooperation of the schools and students who are providing the data. He must be able to store in an organized manner the compositions he has analyzed, for he may find several years after he has begun that there is some new analysis he would like to go back and do or for which he would like to share his raw material with another investigator. Morover, it would be desirable to have someone with enough understanding and interest to carry his work to completion if he cannot.[9]

Instructional Factors Influencing Composition

As the discussion of the Loban study indicated, it is impossible to construct mutually exclusive categories of research studies. That study has dealt with environmental factors, but not to the exclusion of instructional factors. The next study considered here does, however, focus on instructional factors. But one should remember that teachers are part of the student's environment, as is (some say) his own physical makeup—his visual and auditory acuity, his muscular coordination, etc. The student is inextricably intertwined with his environment. Regardless of one's philosophical orientation on the nature of Self, it is convenient at times to categorize things, as is being done here, so long as one is not blinded to the reality that the categories are more a matter of convenience than of fact.

---

[9] At a certain university there is a "roomful of mud" that no one quite knows what to do with. A geologist spent years collecting test borings of the soil of the state. After accumulating thousands of small bags of soil samples, each carefully tagged and packed in the appropriate box with other bags, filling many shelves in a crowded room, the geologist died. Apparently no one at the university understands what analysis the geologist was making or has the interest to complete his work, but no one wishes to discard the data. At a time when space is needed badly on that campus, the roomful of mud sits undisturbed.

An instructional approach designed for retarded college freshmen but, with crowded classrooms, increasingly interesting to high school and college English teachers was examined experimentally by Maize.[9] His experimental group consisted of 75 remedial college freshmen who wrote 26 themes some 250 words in length in addition to a 1,500-3,000 word autobiography, a 3,000-word research paper, and several business and social letters. These papers were read and corrected in class by the students with the guidance of the instructor, who used no time outside of class working with themes. A control group of 74 remedial freshmen wrote fourteen 250-word themes which were marked outside the classroom by the instructor. The control group was taught language usage according to traditional workbook-drill methods. The original study lists the matters covered in the control group. Unfortunately, the pretest and post-test theme examinations are vaguely described, though much attention is devoted in the study to the many objective tests used as well as two scales designed to measure attitudes toward the subject and the instructor. The statistically significant superiority of the experimental over the control group at the end of the study suggests that some combination of frequent writing and student correction is effective. The possibility also is strong, however, that the superiority of the experimental group resulted primarily from the inferiority of the control group, which seems to have been taught a course which many people would now consider outmoded.

As it was one variable in the Maize study, so was frequency of writing a variable in the Buxton study (summarized in Chapter IV), the control group having received no regular instruction and practice in theme writing.[10] Both of these studies and others indicate that college freshmen who do more writing improve their composition more than those who do less writing.[11] A problem which needs investigation, however, is just how much writing is the optimum amount. To carry the problem further than research seems to have done so far, one might well ask *what kinds of writing* following *what kinds of instruction* for *what*

---

[9] Ray C. Maize, "A Study of Two Methods of Teaching English Composition to Retarded College Freshmen" (Unpublished Ph.D. dissertation, Purdue University, 1952). Synopses in *Journal of Educational Psychology*, XLV (January, 1954), 22-28, and, as "A Writing Laboratory for Retarded Students," in *College English*, XVI (October, 1954), 44-48.

[10] Earl W. Buxton, "An Experiment to Test the Effects of Writing Frequency and Guided Practice upon Students' Skill in Written Expression" (Unpublished Ph.D. dissertation, Stanford University, 1958).

[11] See, for one, a study evaluated entirely by objective tests—Virgil L. Lokke and George S. Wykoff's "Double Writing in Freshman Composition: Experiment," *School and Society*, LXVIII (December, 1948), 437-439.

kinds of students. It does not seem reasonable that doubling the number of aimless writing assignments which are then marked in a perfunctory manner would necessarily stimulate students to improve their writing more than developing a somewhat less numerous but carefully planned sequence of assignments which are marked in an instructive manner related to the sequential pattern of the assignments. Another factor which may rival frequency of writing in importance is extensiveness of reading, as the recent Heys experimentation with eleventh graders indicated.[12] But most such studies do not reveal the nature of the writing assignments as clearly as the Buxton study did, and it is thus difficult to compare and contrast one experiment with another. Although most of the research in this area supports the familiar contention that students learn to write by writing, with the help of a qualified teacher, it is not clear yet with what kinds of writing it is true, how much is the optimum amount, and with what kinds of students such generalizations are effective.

Another instructional factor influencing composition learning is the procedure of revision. Is it valuable for a student to revise his papers after his teacher has marked them? Again, the Buxton study affords a tentative "yes" for college freshmen revising their papers under supervision—tentative because revision was one of several variables in that study. It is not clear from Buxton's research, however, which aspects of composition are learned better through revision, if one is concerned with that question. Fellows was, in an experiment he conducted with ninth graders.[13] Although the experiment extended only 12 weeks and the gains were very small, the Fellows study does indicate that a combination of theme correcting by the teacher and revision by the pupil does have a more favorable effect on the grammar and punctuation (but not the spelling or capitalization) of above average ninth graders than does mere writing without correction and revision. In another experiment, involving 1,039 Chicago area children in grades 6 through 9, Lyman demonstrated that even in a period as short as seven weeks pupils can reduce their grammatical and mechanical errors—including spelling and capitalization—more than half by learning how to correct errors before submitting their papers.[14] Remembering that it has yet to be shown what effect revision has on such elements as organization and supporting details, it does seem clear that what little research has been done in this area points out the efficacy of revision for, at least, sixth graders, junior high school pupils, and college freshmen.

---

[12] Frank Heys, Jr., "Report of Experiment in Teaching Composition" (October 15, 1961). Mimeographed. Synopsis, as "The Theme-a-Week Assumption," *English Journal*, LI (May, 1962), 320-322.

[13] John E. Fellows, "The Influence of Theme-Reading and Theme-Correction on Eliminating Technical Errors in the Written Compositions of Ninth Grade Pupils" (Unpublished Ph.D. dissertation, University of Iowa, 1930).

[14] Rollo L. Lyman, "A Co-operative Experiment in Junior High School Composition," *School Review*, XXXIX (December, 1931), 748-757.

It is somewhat surprising that little research has been done in that aspect of a composition teacher's work in which he spends so much of his time—the marking and grading of papers for instructional purposes. (The rating of compositions for purposes of research has been discussed in Chapter II.) Again, the Buxton study seems to offer as good evidence as is available on the relative effectiveness of various grading procedures. One of Buxton's conclusions was that college freshmen whose writing is graded and thoroughly marked improve their writing more than those whose compositions receive a few encouraging general suggestions but no grades or intensive marking. It is not clear, however, to what extent the intensive marking was responsible for that superiority and to what extent revision of the papers merits the credit. Certainly it has not been proved that intensive marking, with or without revision, is the best procedure to use with upper elementary grade or junior high school students.

One study which might have clarified the problem of grading if it had extended beyond its mere 11 weeks is Wormsbecker's thesis.[15] Working with six teachers and 213 sixth graders in three of the Vancouver elementary schools, Wormsbecker investigated the effect on writing improvement of three different grading methods: (1) an overall impression method, (2) a split grade, representing content and form as a "fraction," and (3) a "single point per theme" method, in which the teacher graded each set of themes for a different single consideration each time, the pupils not knowing beforehand which consideration would be used. The fact that there were no significant differences among the effects of the three methods may have resulted from the shortness of the experimental period. A similar investigation was conducted by Halvorson, who compared the effectiveness of indicating errors with symbols and abbreviations to indicating merely the general location of errors by placing check marks in the margins of the theme.[16] Halvorson's results also were inconclusive. In short, research has so far given no clear indication of the most efficacious way to mark papers except for the qualified endorsement by Buxton of thorough marking, a sentence or two of constructive comment at the end, and a grade, and that endorsement does not necessarily hold good for younger people. (A recent review of the studies of various marking procedures may be found in Chapter I of an investigation by McMechan.[17])

---

[15] John H. Wormsbecker, "A Comparative Study of Three Methods of Grading Compositions" (Unpublished M.A. thesis, University of British Columbia, 1955).

[16] Nelius O. Halvorson, "Two Methods of Indicating Errors in Themes," *College English*, II (December, 1940), 277-279.

One of the most heavily investigated problems in the teaching of writing concerns the merits of formal grammar as an instructional aid. Study after study based on objective testing rather than actual writing confirms that instruction in formal grammar has little or no effect on the quality of student composition. Many of the summaries of research listed in Chapter V provide title after title of studies confirming this generalization. For example, Butterfield contrasted the effectiveness of a grammar approach and a "thought" approach in teaching 831 sixth, seventh, and eighth grade pupils to punctuate.[18] He concluded that, although the "students who were taught grammar per se learned significantly more grammar than students who were not taught grammar . . . significantly superior results in punctuation are obtained by direct methods rather than by methods which are based upon a knowledge of so-called related grammatical elements."

Uncommon, however, is carefully conducted research which studies the effect of formal grammar on actual composition over an extended period of time. The Harris study (summarized in Chapter IV) compared the effect of instruction in formal grammar and functional grammar over a period of two years on the writing of 228 London pupils aged 12 to 14.[19] Using a short-answer test on the terms and application of formal grammar as one measure and frequency counts of various types of sentence constructions and common grammatical errors in the students' compositions as a second measure, he concluded that, although the formal grammar classes made significantly higher gains on the grammar test, "the study of English grammatical terminology had a negligible or even a relatively harmful effect upon the correctness of children's writing in the early part of the five Secondary Schools."

In view of the widespread agreement of research studies based upon many types of students and teachers, the conclusion can be stated in strong and unqualified terms: the teaching of formal grammar has a negligible or, because it usually displaces some instruction and practice in actual composition, even a harmful effect on the improvement of writing.

---

[17] Melville Y. McMechan, "The Relative Effectiveness of Four Procedures for Evaluating Students' Written Themes" (Unpublished M.A. thesis, University of British Columbia, 1961).

[18] Clair J. Butterfield, "The Effect of a Knowledge of Certain Grammatical Elements on the Acquisition and Retention of Related Punctuation Skills" (Unpublished Ph.D. dissertation, University of Iowa, 1945).

[19] Harris, *op. cit.*

It may seem incomplete to conclude a discussion of instruction in composition without mentioning some of the valuable descriptions of productive programs and procedures—like the reports of Mearns,[20] Rounds,[21] Hartsig,[22] and Burrows,[23] to mention several studies of creative writing, or the recent analysis by Kitzhaber[24] of college writing programs. Helpful though they are as descriptions and analyses, such investigations have a different approach to truth than the scientific studies which fall within the purview of this report. As a colleague put it, Kitzhaber's book uses a kind of "informed intuition" with its "controls" rooted in the author's accumulated experience.

Rhetorical Considerations

If little has been proved about the instructional factors influencing composition, it is fair to say that almost nothing has been proved in a scientific sense about the rhetorical aspects of written composition. By "rhetorical" is meant here those aspects of writing which (to simplify somewhat) are larger than the unit of the sentence—in expository writing, for instance, the main idea and its analysis; the support of subordinate ideas with details, examples, statistics, and reasons; and the organization of the previous elements into an orderly and meaningful whole. It is a challenge to investigate these aspects of writing in a scientific way.

A rather sweeping attempt to identify some rhetorical considerations as well as other concerns in writing was made by Culpepper-Hagen.[25] After she and two other raters identified the best 30 and the worst 30 writers in a group of 120 freshmen at the University of Denver (purposely excluding the mechanics of expression from their criteria of evaluation), and supplementing the evaluation of three papers from each student by a battery of objective tests and by the student's grade in the Basic Communication course, she administered to the 60 students a Writing

---

[20] Hughes Mearns, *Creative Youth* (New York: Doubleday, Doran, 1925), and *Creative Power* (2nd rev. ed.; New York: Dover Publications, 1959).

[21] Robert W. Rounds, "Teaching Creative Writing in the Junior High School" (Unpublished Ed.D. dissertation, New York University, 1942).

[22] Barbara A. Hartsig, "A Study of the Creative Writing of Children in the Intermediate Grades" (Unpublished Ed.D. dissertation, University of California, Los Angeles, 1956).

[23] Alvina T. Burrows, *et al.*, *They All Want to Write* (rev. ed.; Englewood Cliffs: Prentice-Hall, 1952).

[24] Albert R. Kitzhaber, *Themes, Theories, and Therapy: The Teaching of Writing in College* (New York: McGraw-Hill Book Company, 1963).

[25] Lessie Culpepper-Hagen, "An Investigation of General Factors Relating to the Writing Effectiveness of Freshmen at the University of Denver" (Unpublished Ed.D. dissertation, University of Denver, 1950).

Methods Inventory. Included among her many conclusions were such statements as these about the distinctive tendencies of the best writers:

2. They follow a previously made plan or outline to guide them rather than think out and organize their topic as they write.
3. They have no unusual difficulty concentrating on the job of writing.
4. They feel no handicap because of lack of skill in punctuation, spelling, and grammar; therefore they do not worry about such details.
5. They leave their work largely as they write it, finding little use for elaborate revision and correction by themselves or others.

Although the Culpepper-Hagen investigation offers many leads for further investigation, it suffers, among other things, from the same weakness which plagues most questionnaire studies: one cannot be sure whether a response indicated the student's actual behavior or the behavior which he thought the investigator wanted.

A more focused and behaviorally oriented study on rhetorical considerations is one by Parker.[36] Although not enough details are afforded in this report to enable the reader to evaluate it, the study appears to provide evidence that topic sentences and concluding summaries (in an expository pamphlet 2,000 words in length) increased comprehension for college freshmen. Cross headings and beginning summaries had little or no effect. Although the use of topic sentences increased comprehension when the subjects were tested immediately, no increase was noted when one week had elapsed between reading the pamphlet and taking a test on it.

Some people may say that Parker studied reading, not writing. What he did, however, is what some other investigators should do, too—study the effect of various rhetorical considerations on the reader. This may be a proof-of-the-pudding type of research which has deeper and more far-reaching possibilities than most investigations of writing, which, after all, tend strongly to be investigations of the effectiveness of writing *for composition teachers*. Doubtless many such studies undertaken by specialists in reading may have implications for effective writing, just as similar research by speech scholars (like Goyer[37]) may suggest topics for research in the effectiveness of written composition.

---

[36] John P. Parker, "Some Organizational Variables and Their Effect upon Comprehension," *Journal of Communication*, XII (March, 1962), 27-32.

[37] Robert S. Goyer, "A Study of Individual Differences in Ability and Achievement of College Students in the Organization of Ideas" (Unpublished Ph.D. dissertation, The Ohio State University, 1955).

Objective Tests versus Actual Writing as Measures of Writing

Sometimes treated with more passion than wisdom and often dealt with by people who have a psychological or financial stake in the outcome, the controversy over the relative merits of objective testing and actual writing as measures of writing will not be settled in the comments which follow. This statement may be useful, however, by touching upon some illustrative studies of the problem, bringing into focus some of the major factors involved, and offering some tentative suggestions to help those who have practical decisions to make before more conclusive research has been conducted.

"Objective tests" is usually taken to mean the familiar multiple-choice tests, seldom true-false or completion tests, in measuring the general merit of composition. A variation was the interlinear test developed some ten years ago by the Educational Testing Service. This test attempted to approximate the editing phase of composition by presenting the examinee with a passage which he was to edit in thirty minutes by crossing out incorrect or infelicitous expressions and inserting his attempts at correction and improvement between the lines of the passage. Trained readers, provided with lists of preevaluated possible responses for each of the problem situations consciously introduced into the passage, assigned each examinee's response a number of points, depending upon which of the listed possibilities the examinee's response most closely approximated. Although a test reliability of .81 and a reader reliability of .95 to .96 were reported for the interlinear test,[28] the Educational Testing Service has sharply reduced its use of it. One can conjecture why; examinees' responses would not often enough match the lists of "possible" responses, and the readers, even though trained, would disagree on their evaluations of responses not anticipated in the tests. Fretting over the examinees' human departures from the anticipated responses, the examiners would find it difficult to resist using the anticipated responses as the possible answers in a multiple-choice form of the same test. In addition, the interlinear test doubtless is characterized by the problems common to the more usual multiple-choice tests while much more expensive to score.

In general, multiple-choice tests of composition suffer from a number of shortcomings. The most common objection heard among composition teachers is that they do not require the examinee to select his own words and to compose--to formulate and organize his own ideas into paragraphs and sentences; such tests are therefore, say these critics, inevitably not valid measuring instruments.[29] The defenders of multiple-choice tests are quick to respond; perhaps even acknowledging what may seem to be "self-evident invalidity," they point nevertheless to the helpfulness of these objective measures as predictors of success in future composition classes[30] and to the notorious unreliability of the grading of actual compositions.

---

[28] Frances Swineford, "Reliability of an Interlinear Test of Writing Ability," *School and Society*, LXXXI (January 22, 1955), 25-27.

Certainly it is true that disagreement among theme graders is common. Perhaps the clearest testimony of this unreliability was developed in a study by Diederich and two colleagues.[31] They analyzed the way ten English teachers rated 300 two-hour compositions by college freshmen in comparison to 43 other raters: social scientists, natural scientists, writers and editors, lawyers, and business executives. The raters were given no standards or criteria for judging the papers, merely asked to sort the themes into nine piles in order of general merit, with not less than 4 percent of the papers in any pile. It was "disturbing to find that 94 percent of the papers received either seven, eight, or nine of the nine possible grades, that no paper received less than five different grades, and that the median correlation between readers was .31. Readers in each field, however, agreed slightly better with the English teachers than with one another." Not enough better was the reliability reported by Noyes for the one-hour English Composition test included among the Achievement Tests used by the College Entrance Examination Board during 1943 and 1944: "A lot of people worked very hard at setting and scoring the themes, in six successive tests during 1943 and 1944. The highest reliability of reading we ever attained was about .57. . . ."[32] Many similar instances of unreliable grading have been reported by other investigators.

But the defenders of objective tests of writing ability seem to overlook or to regard as suspicious the high reliabilities obtained by some investigators. Noyes himself refers to reliability indices of .87, .88, and .87 attained with the one-hour theme which was part of the CEEB

---

[29] Earle G. Eley wrote "An adequate essay test of writing is valid by definition; that is to say, it has face validity since it requires the candidate to perform the actual behavior which is being measured" in "Should the General Composition Test Be Continued?" *College Board Review*, No. 25 (Winter, 1955), 11.

[30] Here is the way it was put by Edward S. Noyes, Vice President of the College Entrance Examination Board, in "Teaching and Testing of English," *College Composition and Communication*, XII (February, 1961), 37: ". . . English is a combination of skills and knowledge, so intricately intertwined that a direct measurement of one is likely to afford an indirect measure of another."

[31] Paul B. Diederich, John W. French, and Sydell T. Carlton, *Factors in Judgments of Writing Ability*, Research Bulletin RB-61-15 (Princeton: Educational Testing Service, 1961).

[32] Noyes, *op. cit.*, p. 35. In a recent article, however, Noyes reports a reading reliability coefficient of .92. See "Essay and Objective Tests in English," *College Board Review*, No. 49 (Winter, 1963), 7-10.

Comprehensive Examination in English before World War II. In studies summarized at length in Chapter IV, Buxton reported reader reliability coefficients of .91 and .88, and Kincaid obtained reliabilities ranging from .77 to .91. In one of the most thorough explorations of this problem, Finlayson found reliabilities ranging from .786 to .961 in the rating of one-hour papers written by 850 sixth graders in 21 Edinburgh primary schools.[33] It seems clear, then, that high reader reliabilities are possible. Suggestions were made in Chapter II for obtaining such reliable grading.

The most serious charge against multiple-choice tests of writing is their lack of validity. Not only do they not require the examinee to perform the actual behavior being measured—he does no actual writing; but these tests also make little or no attempt to measure the "larger elements" of composition, even indirectly. In her contrast of a multiple-choice test and the rating of actual writing, Huddleston noted that the objective test contained 27 punctuation items, 33 problems in idiom, 47 in grammar, and 70 in sentence structure—evidently no items on the clarity, focus, and analysis of the main idea; the logic and clarity of the organization; or the soundness and specificity of the supporting content. It is little wonder, then, that the written part of her test consisted of three 20-minute essays; she did not have to be concerned with affording the examinees time to analyze and formulate their ideas.[34] Although his analysis, too, was made of a test which (though widely used at the time) has been largely supplanted now, Vordenberg found invalidity even within the "smaller" elements of composition, which the objective test attempted to cover. An analysis of the errors in freshman themes in comparison to the types of errors included in the test led him to conclude that capitalization errors were emphasized more on the test than their occurrence in writing warranted, that the seriousness of sentence fragments in writing was not reflected in the weight they received in the test, and that other types of errors were insufficiently represented, if at all, in the test: wordiness and repetition, omissions, superfluous commas, italics, various hyphen problems, certain types of spelling errors, etc.[35] And, again, the examination did not test for such matters as unity, organization, and content. Defenders of the validity of multiple-choice composition tests respond by pointing to the "coverage" of such examinations; the examination may insure that the examinee faces problems which he might avoid if he were doing his own writing. This defense remains to be proved.

---

[33] Douglas S. Finlayson, "The Reliability of the Marking of Essays," *British Journal of Educational Psychology*, XXI (June, 1951), 126-134.

[34] Edith M. Huddleston, "Measurement of Writing Ability at the College-Entrance Level: Objective vs. Subjective Testing Techniques," *Journal of Experimental Education*, XXII (March, 1954), 165-213.

[35] Wesley P. Vordenberg, "A Comparison of Errors Made by College Freshmen on the Cooperative English Test, Form PM, and in Themes" (Unpublished Ph.D. dissertation, Indiana University, 1950). Synopsis, as "How Valid Are Objective English Tests?" *English Journal*, XLI (October, 1952), 428-429. For a now outdated but more comprehensive study of this type, see Matthew H. Willing's *Valid Diagnosis in High School Composition*, Contributions to Education, No. 230 (New York: Bureau of Publications, Teachers College, Columbia University, 1926). Synopsis, as "Individual Diagnosis in Written Composition," *Journal of Educational Research*, XIII (February, 1926), 77-89.

A much more soundly based and frequently substantiated contention of the defenders of objective tests is that they can predict success in composition study more reliably than can the defenders of examination by actual writing. One study substantiating this superiority was conducted in a freshman English course by Stalnaker in 1932. Among 1,500 freshmen who had taken a placement examination consisting of "several short samples of writing of one or two paragraphs each" as well as an objective test, he identified 38 cases in which the student had scored above the 75th percentile rank on one test and below the 25th percentile rank on the other. Of the 38, there were 23 students about whom information was available, and Stalnaker asked their instructors how well the students had done in their composition classes. Stalnaker concluded, ". . . in the case of marked differences between the essay and objective ratings of a student in English composition the objective rating is probably the more valid estimate. Students who receive high essay scores and low objective scores tend to be an inferior group as judged on the basis of their intelligence-test scores or on the basis of their class performance in English composition."[36] Certainly one has reservations about this study. The number of cases used for this conclusion is small, and the shortness of the writing samples probably precludes opportunity for the freshmen to demonstrate well their ability to handle the larger elements of composition. Furthermore, one may ask questions about the reliability of the essay grading and the emphasis of the composition course. In how many schools has objective testing been a good predictor of success in composition classes precisely because those classes have emphasized grammatical and mechanical matters with little or no emphasis on central idea, analysis, organization, and content?

Doubtless one reason that the College Entrance Examination Board has given up grading the composition it includes in its examination is that colleges differ in their emphases in composition instruction. How can the CEEB rate papers in a manner valid for all colleges if some colleges emphasize narration, others exposition, still others argument or literary criticism? And within those emphases, one college may stress content and organization, another style, still another "correctness." The CEEB has found a partial solution by requiring a paper on an assigned topic and forwarding the paper to the candidate's preferred institution for evaluation. But even that is not completely satisfactory, for doubtless a number of colleges would prefer different topics, eliciting the kind of writing they attempt to teach. Until that dull, gray utopian day when all freshman composition courses have the same emphasis, the CEEB composition examination will have varying validities for the institutions receiving its papers. Meanwhile, the CEEB continues to include an essay examination in its battery of tests, partly to counteract the critics who fear a "backwash effect" on the secondary schools if no actual writing is demanded.[27] The CEEB probably discounts this argument, however, for French discovered from questionnaires returned by 189 out of 224 high school English teachers that, although the respondents "were overwhelmingly in favor of essay testing," the great majority seemed not to be aware of whether or not the CEEB examinations included such a test.[28]

---

[36] John M. Stalnaker, "Essay and Objective Writing Tests," *English Journal* (College Edition), XXII (March, 1933), 222.

In view of all these conflicting beliefs, what is the English teacher, department head, or chairman of freshman composition to do about objective tests and actual writing as measures of writing ability? The answer must depend in part upon what he wishes to accomplish.

If a school department head or college chairman of freshman composition wishes to section large numbers of students according to their predicted success in composition, he probably will find a relatively sound objective test the most efficient instrument. Although some mis-sectioning doubtless will occur, usually the individual students can be reclassified during the first week or two of the year on the recommendation of their teachers. (This would not work, of course, with the 11+ examinations used in England. There a consideration of actual writing seems in order.) The expense and effort involved in reliable rating of actual compositions are not often justified where only a rough sorting of large numbers is needed and where occasional errors in classification can be corrected with little pain. Meanwhile, those who construct such objective tests should continually strive to improve and update them.[29] Especially they should bend their best efforts to include items on the larger elements of composition.

---

[27] See, for instance, the articles of Stephen Wiseman, and Edward S. Noyes' "Essay and Objective Tests in English," *College Board Review*, No. 49 (Winter, 1963), 10.

[28] John W. French, "What English Teachers Think of Essay Testing," *English Journal*, XLVI (April, 1957), 196-201.

[29] Some excellent suggestions are offered by James B. Enochs in "Measuring the Ability to Write" (Unpublished Ph.D. dissertation, University of Chicago, 1948).

If a teacher is interested in diagnosing the strengths and weaknesses in various aspects of the composition of individual students, he will do best to base his diagnosis on an examination of their actual writing, realizing that he must gear the assignment to the matters he wishes to diagnose. The objective testmakers seldom claim that their tests are useful for this purpose, and those who do have yet to substantiate their claims. When a beginning teacher does not feel confident of his ability to diagnose his students' writing, he would be wise to exchange a set of themes for a set from a more experienced colleague and then study the colleague's comments.

If a teacher or chairman wishes to evaluate the writing performance of students after instruction, he should plan carefully to use several papers written by the students. Objective tests will not accomplish this purpose; if the teacher has been emphasizing rhetorical matters as well as grammar and mechanics, objective tests simply are not valid—they make little or no attempt to measure the larger elements of composition. But, as Kincaid demonstrated (see Chapter IV), one composition is not enough, especially for more capable students.[40] If a large number of students is to be evaluated on a pass or fail basis, those students who do not pass a first composition may be given a second opportunity; there is little point in examining a second time the student who passes the first. But if students are to be evaluated for the purpose of assigning grades, probably they should be given as many as three compositions, with only the best two of the three grades being averaged for the final grade. The assignments and the basis of evaluation should be similar to or the same as those used in instruction, and the papers should be graded by at least two raters using a mutually understood and accepted rating system with which they have previously practiced. Other considerations were discussed in Chapter II. Additional suggestions may be found in articles by Diederich,[41] Gerber,[42] and Wiseman.[43]

Other Considerations

Size of English Classes

Although there has been much research on size of class in general, little has appeared on the size of English classes and less on its influence on written composition. The Smith study, published in 1931 and summarized in Chapter IV, investigated the effect of class size on various aspects of ninth grade English.[44] It seems possible from that investigation that teaching composition to superior ninth graders may be generally as effective in large classes as in small classes if the teacher has the requisite ability and is given adequate time to prepare and adequate assistance in handling the routine aspects of class management. Now that the overhead projector has been tried out with reported success in a number of large lecture situations, it seems time for a new experiment in size of class, using lay readers and the overhead projector as aids to composition instruction.

---

[40] Gerald L. Kincaid, "Some Factors Affecting Variations in the Quality of Students' Writing" (Unpublished Ed.D. dissertation, Michigan State College, 1953).

[41] Paul B. Diederich, "The Measurement of Skill in Writing," *School Review*, LIV (December, 1946), 584-592.

[42] John C. Gerber, "Testing and Evaluating in the Skills of Communication," *College English*, IX (April, 1948), 373-384.

[43] Stephen Wiseman, "The Marking of English Composition in Grammar School Selection," *British Journal of Educational Psychology*, XIX (November, 1949), 200-209.

Lay Readers

One way in which more than routine assistance can be given the English teacher is to provide the services of a "lay reader," usually a person not certified to teach in the public schools but qualified by other experience to assist in the marking and grading of compositions. No convincing study of the effectiveness of a lay reader program was discovered by those preparing this report, and that is not surprising. It is a natural thing—as in the Sauer study[45]—to have a lay reader experiment in which the experimental and control classes are taught by the same teacher, automatically alleviating the *total* instructional load of the teacher, who may tend, then, to devote more time to his unassisted control class because he is being assisted by the lay reader in his experimental class. It seems clear that a carefully designed experiment with lay readers necessitates, in part, a time study of the teachers involved. Another unique necessity, if the findings are to have meaning to school systems outside the experiment, is to include some objective descriptions of the qualifications of the lay readers and teachers involved; otherwise, a conclusion valid for a high school in, for example, a university community where there are many competent faculty wives who would like to be lay readers has little meaning for a less fortunate community where there is no abundant supply, if any, of people competent to mark compositions. If one has read the Dusel report and is cognizant of the enrollment pressures on public schools and the shortage of first-rate

---

[44] Dora V. Smith, *Class Size in High School English: Methods and Results* (Minneapolis: University of Minnesota Press, 1931).

[45] Edwin H. Sauer, *Contract Correcting: The Use of Lay Readers in the High School Composition Program* (Report to the Cooperative Research Program of the United States Department of Health, Education, and Welfare, rev. ed., unpaged, 1962).

English teachers, he knows that the use of lay readers deserves careful investigation.[46]

Teaching by Television

Despite the fact that much money has been granted by foundations for experiments in the use of television as an instructional aid, little of the research, as published, seems convincing, at least where instruction in written composition is concerned. Often the nature of the course of instruction is not made clear, or the evaluation of the experiment is based solely on objective testing, or the attitudes of the students have not been carefully examined. In one experiment, the cost of the equipment and studio technicians seemed to exceed by far the salaries which were saved by reducing the number of instructors. Sometimes glowing published reports seem to overlook rumors of inattentive students and disgruntled teachers whose normal classroom procedures have been interrupted rather than supplemented by the television set in the front of the darkened room. Until such time as more conclusive research has been published, one is tempted to reserve the television teaching of composition for correspondence students who cannot be present on a college campus.

The above criticisms do not seem very applicable to the Becker study (summarized in Chapter IV), in which kinescopes (films made with television cameras) were used in the teaching of college composition.[47] Kinescopes, like teaching films, can be used with more flexibility of timing than television, even though they were not so used in the Becker experiment. Although the kinescopes were not always technically smooth and the results of the experiment would have been more convincing if a trial period had permitted the instructors to adapt their teaching better to the medium being used, the study does indicate that kinescopes may be a useful instructional aid, especially where instructors are teaching composition for the first time. The study makes much the same claim for releasing freshmen from one class meeting each week (of four usual meetings) to supplement their classwork by library study guided by a bibliography of readings on, among other things, rhetoric and written composition. After reading the Becker study, one wonders if a series of professionally prepared and mature teaching films on composition would not prove even more effective with freshmen than kinescopes or television. The problem is to find teaching films which are well done and which cover the particular matters emphasized in the individual school or college.

---

[46] William J. Dusel, "Determining an Efficient Teaching Load in English," *Illinois English Bulletin*, XLIII (October, 1955), 1-19. (This issue of the Illinois bulletin is distributed by the National Council of Teachers of English.)

[47] Samuel L. Becker, *et al.*, *Communication Skills: An Experiment in Instructional Methods* (Iowa City: University of Iowa, 1958).

Writing Vocabulary

A fruitful product of research on writing vocabulary has been the word count—a count of the frequency with which words appear in various kinds of writing. Most apparently pertinent to this pamphlet are the counts based on the writing of students and adult nonprofessional writers. Not designed for spelling lists, they are guides to persons wondering which words may be used commonly enough at a given grade level to warrant inclusion in a spelling list for that grade. Word lists based on children's writing certainly are not designed to be used for reading vocabulary study; reading vocabulary begins to exceed writing vocabulary soon after a child starts to read and quickly becomes far more extensive than his writing vocabulary. Graded writing vocabulary word lists can also be useful guides in judging one aspect of the maturity of a pupil's writing. Among the more carefully constructed of these word lists are the following, all of which antedate any impact of television on vocabulary:

Buckingham, Burdette R., and E. W. Dolch. *A Combined Word List.* Boston: Ginn and Company, 1936. 185p.

The sources of this list are (1) lists of words which children were asked to construct in grades 2 through 8 in Illinois, Manhattan, and Brooklyn and in grade 2 in New England, and (2) ten other word lists of various kinds, including those based on children's writing, speech, and literature. The sources and recommended grade levels are indicated after 10,000 words of the 19,000.

Fitzgerald, James A. *A Basic Life Spelling Vocabulary.* Milwaukee: Bruce Publishing Company, 1951. 161p.

This list of 2,650 words was selected from a much larger list compiled through study of seven other lists drawn from children's writing, both in and out of school, and from adult writing.

Horn, Ernest. *A Basic Writing Vocabulary: 10,000 Words Most Commonly Used in Writing.* University of Iowa *Monographs in Education.* Iowa City: College of Education, University of Iowa, 1926. 225p.

The sources of this list are business correspondence, personal letters, and a variety of other letters and types of writing, written by college students and adults. Although this list was published in 1926, many people still consult it because of the diversity and level of its sources.

Rinsland, Henry D. *A Basic Vocabulary of Elementary School Children.* New York: Macmillan Company, 1945. 636p.

This list is constructed from the writing of children in grades 1 through 8, submitted by the officials of 708 schools throughout the United States. The graded list includes 14,571 words.

Thorndike, Edward L., and Irving Lorge. *The Teacher's Word Book of 30,000 Words.* New York: Bureau of Publications, Teachers College, Columbia University, 1944. 274p.

Based on the Thorndike general count of 1931, the Lorge magazine count, the Thorndike count of 120 juvenile books, and the Lorge-Thorndike semantic count, this list gives the frequency of occurrence of each word according to each of the four sources and a frequency which represents a summary from all four counts. Although based on professional writing, this list is still useful in developing criteria to judge the maturity and style of student writing or to guide the selection of words by the professional writer who does not wish his vocabulary to be too difficult for his prospective readers.

Other types of research in writing vocabulary would be valuable, too—studies in the range, freshness, and precision of diction at various levels of maturity and intelligence, and studies of the variation in the connotations which words have for readers of different types. A promising procedure for getting at the range of a student's writing vocabulary was used in a study by Newman.[illegible] Not having his subjects actually write sentences or paragraphs, he gave them passages in which a number of spaces were left blank, in each of which the students were to write a different appropriate synonym of the master word provided. From determining how many of the blanks a student was able to fill, he gauged the range of the student's expressional vocabulary.

Spelling

Research in spelling is barely touched upon in this report for the simple reason that, however important accurate spelling may be in the

---

[illegible] Herbert M. Newman, "A Study of Range in Expressional Vocabulary" (Unpublished Ph.D. dissertation, Columbia University, 1930).

clarity and social acceptability of composition, many of the factors of good spelling do not seem to be closely involved with the factors of good composition. A number of investigators have shown, for example, that it is more effective to base the selection of words for spelling instruction for children on carefully prepared spelling lists than to base selection on spelling words chosen from the compositions of the pupils. A recent confirming study of this type was conducted by Deacon, working with 18 teachers and 414 second and third graders.[49]

One study of spelling which should be on every composition teacher's reference shelf, although it is not a study of spelling in student writing, is the Emery report.[50] Emery compared the variant spellings of the same words in five standard college-level dictionaries and found many discrepancies, suggesting that teachers should exercise caution (by consulting the 28-page "Reference List of Variant Spellings" in the Emery pamphlet) before marking some words as misspelled.

One final word regarding spelling research is in order. Although it evidently makes sense to teach children in grades 1 through 8 to spell a basic core of words from high-frequency lists rather than from words they misspell in their own compositions, that is no reason why studies of misspelling should not be based on actual misspellings in student compositions rather than on multiple-choice tests or dictated lists. When one reads published reports on misspelling, the printing of the report obscures a great truth—that many words in student writing have been formed so carelessly that one cannot be sure whether or not they are misspelled, or, if so, how. More use of photographs of handwritten words in published research would help to convey some sense of the problem the spelling investigator faces in illegibility.

Handwriting

Like spelling, handwriting appears to be a field of research which more often than not has less to do with composition than one might suppose. As with spelling, the writers of this report made no attempt to review the thousands of published reports on the subject, although they did read those few studies recommended to them and listed in the bibliography in Chapter V. Two observations do seem pertinent. An alarming proportion of the small sample of studies read seems to be based on the handwriting used when pupils copy some sentences from the blackboard or, more commonly, write from memory some familiar passage like the opening lines of the Gettysburg Address. Surely such artificial circumstances cannot yield a valid sample of pupils' typical handwriting, done during the heat of composition. The other observation is more in the nature of a request—that the titles of reports on handwriting use the word "handwriting," not the ambiguous "writing."

---

[49] Lawrence S. Deacon, "A Comparison of Two Programs of Instruction in Spelling" (Unpublished Ed.D. dissertation, University of California, Berkeley, 1953).

[50] Donald W. Emery, *Variant Spellings in Modern American Dictionaries* (Champaign, Illinois: National Council of Teachers of English, 1958).

Typewriting

The effect of typewriting on progress in written composition evidently has been the subject of very little investigation, most of that with elementary school children. Moreover, because most, if not all, of the studies have been subsidized by grants from typewriter manufacturers, one cannot help but feel some reservations about the findings. One such investigation, conducted in the late twenties by Wood and Freeman, and utilizing 6,125 pupils in the typewriter group, 8,824 in the control group, concluded that typewriting stimulated more and longer original compositions, both in typewriting and in handwriting, especially at the first and second grade levels.[11] The most favorable influence of typewriting seemed to be on spelling, as measured by that portion of the Stanford Achievement Test. Generally speaking, Wood and Freeman found the typewriter equally helpful to bright, normal, and dull pupils. A more recent study by Erickson and Clow, based on 24 matched pairs of fifth graders, showed mixed results.[12] Certainly the effect of the typewriter on composition deserves careful investigation, financed by sources other than manufacturers. The ease with which kindergarten and primary school children seem to be able to operate portable electric typewriters argues especially for a careful investigation in that area, perhaps using Van Bruggen's "rate of flow" index, described early in this chapter, to determine whether or not the typewriter can increase the fluency of writing in general.

Relationships of Oral and Written Composition

With the realization revivified by structural linguists that written language evolves from oral language, the need is more apparent than ever for investigations into the rhetorical and syntactic relationships between the two types of language. The Harrell study did investigate some features of the relationships in the writing and speaking of children in grades 4, 6, 8, and 10.[53] His language samples seem artificial, however, inasmuch as the compositions were merely narrative reproductions of a story and the oral narratives were delivered into a visible microphone when the speaker was alone with the investigator. On the other hand, Harrell's idea of using ten-minute films to trigger oral and written composition appears to be a good means for achieving standard stimuli. If the content of films—especially silent films, to forestall copying—can be tied to pupils' desires to write, perhaps more original, spontaneous language samples will result. The relationships between oral and written composition are still wide open for research, as is the important matter of stimulating a desire to write well.

---

[11] Ben D. Wood and Frank N. Freeman, *An Experimental Study of the Educational Influences of the Typewriter in the Elementary School Classroom* (New York: Macmillan Company, 1932).

[12] Lawrence W. Erickson and Chton Clow, *The Manual Portable Typewriter as a Tool of Learning with Fifth-Grade Elementary School Pupils* (New York: Teachers College, Columbia University, 1959).

Unexplored Territory

Some questions which seem fundamental in the teaching and learning of written composition apparently have gone almost untouched by careful research. This chapter concludes with a list of questions not considered previously in this chapter, which indicate areas in which future investigators may wish to direct their efforts:

1. What kinds of situations and assignments at various levels of schooling stimulate a desire to write well?
2. What do different kinds of students prefer to write about when relieved of the expectations and requirements of teachers and others?
3. What are the sources of fear and resentment of writing?
4. How do the kinds of writing which adults compose vary with their occupations and other factors?
5. What is the effect on writing of having the student compose his paper for different kinds of readers?
6. At which levels of maturation does it seem appropriate to introduce the various modes of discourse—narration, poetry, drama, exposition, argument, and criticism?
7. What is the relative effectiveness of writing shorter and longer papers at various levels of maturity and proficiency?
8. At which levels of maturation does it seem appropriate to introduce the various rhetorical elements of writing?

---

[53] Lester E. Harrell, Jr., "An Inter-Comparison of the Quality and Rate of the Development of Oral and Written Language in Children" (Unpublished Ph.D. dissertation, University of Minnesota, 1956). Published in *Monographs of the Society for Research in Child Development*, XXII, 3 (1957).

2

9. What are the effects of various kinds and amounts of reading on the quality and kinds of writing a person does?
10. What are the direct and indirect effects of particular sensory experiences and guided observation upon writing?
11. At what stages of maturity do students spontaneously *seek* specific help in improving particular aspects of writing, such as specificity of details, transitions, parallel structure, and metaphor?
12. At which levels of maturation *can* particular aspects of writing most efficiently be learned?
13. Does the oral reading of rough drafts help the elementary school child strengthen "sentence sense"? How does it?
14. What techniques of composition most effectively help build self-discipline and pride in clarity, originality, and good form?
15. What procedures of teaching and learning composition are most effective for pupils of low socioeconomic patterns?
16. What procedures of teaching and learning composition are most effective for pupils learning to write English as a second language?
17. Can study of the newer types of linguistics help writers?
18. Can formal study of rhetorical theory or of logic help writers?
19. How is writing affected by extensive study and imitation or parody of models?
20. What forms of discourse have the greatest effect on other types of writing? For example, does writing poetry help a writer of reports?
21. What is involved in the act of writing?
22. How does a person go about starting a paper? What questions must he answer for himself?
23. How does a writer generate sentences?
24. Of what does skill in writing really consist?

PRECEDING PAGE BLANK- NOT FILMED

IV.

SUMMARIES OF SELECTED RESEARCH

Basis for Selecting These Studies

One strong impression which should emanate from Chapters II and III is that there are many variables to control in research with written composition. There are so many that it is an unusual study which does not leave several important variables uncontrolled or undescribed. Consequently, the writers of this report were tempted more than once not to select any "most soundly based studies" to summarize at length, as the Executive Committee of the National Council had requested. (See Chapter I.) Recognizing, however, that "most soundly based" does not mean "perfect in all respects" but "most soundly based of all those studies available," the writers have selected five distinctly superior investigations to summarize here in Chapter IV. Because three of these studies have never been published and the other two are out of print, the summaries are detailed. The fullness of the summaries serves to inform the reader who wishes to determine for himself whether or not to be convinced by the conclusions of the studies, the investigator who wishes to replicate or extend the experiments described, and the student of research who wishes to discover how some superior investigations have been designed, executed, and reported.

Although the methods of research suggested in Chapter II are, in effect, criteria which can be used in selecting "most soundly based studies," the details of those suggested methods (or criteria) tended to become apparent to the writers of this report as they screened the studies recommended to them. From the beginning of this screening, however, certain broad criteria were employed. Generally, the selected studies, the writers agreed, should do these things:

1. Base the investigation at least in part on the direct observation of actual writing, not entirely or mostly on such indirect measures as objective tests or questionnaires.
2. Study the writing either of a generous number of students (never actually specified, "generous" seemed to mean at least

seventy or eighty) or of as few as twenty students who were very carefully selected or very carefully matched with another twenty.

3. Describe the procedures in a controlled experiment or the features of writing in a textual analysis in enough detail that it is clear what was being studied.
4. Use procedures of statistical analysis which, though not necessarily complicated, were appropriate and consistent and did not obscure the raw data being analyzed.
5. Maintain as objective an investigation as possible by controlling or reporting the salient variables; that is, by keeping the investigator as "removed" from the study as possible, by preserving the anonymity of the students when evaluating or analyzing their writing, by describing the abilities of the pupils used, etc.

As the writers of this report screened the studies to a more and more limited number, they naturally became increasingly detailed in their expectations, drawing more heavily from the considerations covered in Chapter II. They sometimes had to weigh an outstanding merit against a striking flaw or to consider the desirability of representing a particular type of study in this report. Moreover, they anticipated some criticism because one of the five studies had been conducted by a member of the ad hoc committee (the Smith study) and another had been accomplished at the institution with which the three writers of this report are associated (the Becker study). But the writers finally concluded, with the support of the ad hoc committee, that those two investigations should be included because of their intrinsic merits and interest, despite the charges of favoritism which might be made.

Explanation of Statistical Terms

For the benefit of the reader who is not familiar with statistical terminology, the writers include here explanations of the terms used in a number of the reports.

Level of confidence

Because there are always a number of uncontrolled factors operating in an experiment, the effects that are observed cannot all be attributed to the particular treatment under consideration. There are available a number of statistical procedures to determine the probability that an observed effect is due to these "chance" factors rather than to the experimental variable. This probability is referred to as the level of confidence or significance; for example, if an effect is significant at the .01 level of significance, it means that there is only one chance in one hundred that differences as large as those observed could have occurred by chance.

Variance

The spread or scatter among the scores in a distribution is frequently a very useful bit of information to have. One such measure of scatter is the variance, the square root of which is referred to as the standard deviation. In addition to their descriptive usefulness as measures of variability, the variance and standard deviation are used in a number of the statistical tests that are employed to determine significance of an effect.

Analysis of variance

This is a statistical procedure frequently used to test the significance of an effect, particularly when there are more than two experimental groups. Basically it involves comparing different estimates of the variance of a distribution. These comparisons yield an F-value which when used with appropriate tables indicates the level of the significance of an effect.

Analysis of covariance

It is frequently desirable to match subjects in an experiment, but this matching becomes very difficult when dealing with certain types of subjects, e.g., intact classrooms. Analysis of covariance is a statistical procedure which adjusts the scores obtained in such a way as to match the groups statistically when actual matching is not possible.

Chi-square

Chi-square (χ^2) is used when there is a need to determine whether or not a distribution of scores is like some hypothetical distribution the experimenter has set up. A value based on the discrepancies between the actual and hypothetical scores is referred to a table, and significance of the discrepancies can then be determined.

Critical ratio

Most tests of significance are based on a ratio of some sort. The ratio varies with the size of the differences between or among the means. The critical ratio is that ratio which is of such a magnitude as to be considered statistically significant.

F-ratio

The F-ratio is a statistic that is used under certain conditions to test whether or not the results of an experiment may be considered to have arisen by chance.

t-test

The t-test, like the F-ratio, is a statistic used to determine the significance of an effect.

At the beginning of each summary, a bibliographic reference to the study is given. When the University Microfilm number is included, the dissertation cannot be borrowed through interlibrary loan but must be purchased, in microfilm, directly from University Microfilms, 313 North First Street, Ann Arbor, Michigan.

The Buxton Study

Buxton, Earl W. "An Experiment to Test the Effects of Writing Frequency and Guided Practice upon Students' Skill in Written Expression." Unpublished Ph.D. dissertation, Stanford University, 1958. 135p. University Microfilm No. 58-3596. A synopsis of this study has been published in the *Alberta Journal of Educational Research,* V (June, 1959), 91-99.

Purpose

The purposes of this experiment, conducted at the University of Alberta during the academic year 1956-1957, were to determine (1) whether or not "regular practice in writing over a period of seven months (a University of Alberta 'year') [would] result in a significant improvement in writing skill" and (2) which of the following methods was superior in improving writing skill: (a) a "freedom from restraint," or "Writing," method in which no interlinear or marginal marks were employed on the student papers, no grades were noted on the papers, the only remark was a paragraph of generous comment at the end of the paper, and the students were not advised to correct or revise their papers, and (b) a "prevision, writing, and revision," or "Revision," method, in which the papers were thoroughly marked, graded, commented upon (adversely, when necessary) in a paragraph at the end of the paper, and discussed by the raters and revised by the students during 35 to 50 minutes of class time when each paper was returned. Thus the experiment attempted to test, at the college freshman level, the effectiveness of two divergent views toward the teaching of composition.

The reader's attention is directed to the unusual opportunity provided at Alberta for minimizing the effects of factors external to an experiment: all 257 subjects were freshmen registered in the Junior Elementary Program, a "one-year 'emergency' course designed to train teachers for Alberta schools." All of these students took the same university courses (though not all taught by the same instructors), spent the same number of hours in class, and wrote the same midterm and final examinations. The entire class of students was divided into six groups of equal size, each group taking all of its classes together throughout the year.

Procedures

For the experiment, the 257 students were each assigned a number from a table of random numbers, and then the names were arranged in numerical order and the new list divided into six equal classes: the Control group (Classes A and D) with 86 students, the Writing group (Classes B and E) with 86 students, and the Revision group (Classes C and F) with 85 students.

To minimize different influences on the three groups as a result of the varying personalities and teaching techniques of different instructors, the investigator arranged, "as nearly as possible, that no two classes in the same group were taught by the same instructor unless the instructor taught more than three classes," in which case he taught a class in each group. To show how this was done, Table 1 and its accompanying notes are copied here from pages 23 and 24 of the original report.

The three groups of students received distinctive instruction for the purposes of the experiment.

The Control group took the subjects indicated in Table 1, doing the usual assignments which those courses required but not doing the writing required of the Writing and Revision groups, as explained below.

In addition to the courses taken by the students in the Control group, the students in the Writing and the Revision groups wrote one essay of approximately 500 words each week for a total of sixteen papers. Each week the instructor of a different course assigned the essay topic for that week, the students writing one paper for each course each semester. A schedule was distributed to all of the students, instructors, and readers which indicated that an assignment was given toward the end of each week, was submitted by the student to the instructor on the first

Table 1

Courses and Instructors in the Junior Elementary Program at the University of Alberta

Course Number	Name	Class Hours	Instructors					
			Control Group		Writing Group		Revision Group	
			A	D	B	E	C	F
Ed.101	Physical Education	48	5	7	5	7	5	7
Ed.105	Health	48	12	7	12	7	12	7
Ed.106	Social Studies	72	Social Studies Staff					
Ed.121	Primary Work	72	14	14	3	3	3	14
Ed.127	Mathematics	36	10	10	10	13	13	13
Ed.129	Science	48	17	17	17	17	17	17
Ed.138	English	72	20	2	18	6	2	18
Ed.151	Art	48	21	8	8	21	21	8
Ed.153	Music	48	21	1	1	21	21	1
Ed.161	Administration	36	19	19	22	22	23	23
Ed.176	Psychology	72	11	15	9	9	11	15

Notes on Table

1. Numbers in columns under Group C, Group W, and Group R refer to instructors; thus, Affleck is number 1, Buxton number 2, Caldwell number 3, and Worth number 23.
2. Education 106 (social studies) was taught by instructors 4, 16, and 23. Each instructor taught one unit to all classes; therefore, each class had the same allotment of time with each instructor.
3. The division of classes among instructors in Education 121, Education 127, and Education 161 was not so desirable as it might have been for this investigation. However, since the assignments of these classes largely involved collecting materials for primary work, surveying primary texts, working with numbers and number materials, observing classes, and planning timetables, it is felt that any differences in students' experience in essay writing were negligible.

day of class in the next week, and was returned to the students during the first day of class in the following week.

The treatment of assignments and papers in the Writing group was characterized by "freedom from restraint." If a student did not wish to write on the topic assigned, he was encouraged to choose any topic he desired, even an anecdote, imaginative experience, or personal happening. No interlinear or marginal marks or comments were written on their papers, but a three- or four-sentence comment was written at the end, giving a general impression of the essay by praising it as much as possible but also by suggesting one or two ways in which the next paper could be made clearer or more interesting. No grades of any kind were placed on the Writing group essays; they were returned to the students without comment, and no suggestion was made that the papers be revised or corrected.

The assignments and papers in the Revision group were treated with considerably more direction. Although the students could develop the assigned topic "in their own way," they were expected to write on the same topic and to include some critical thinking, a central idea, and material that was organized and developed. They were encouraged to organize preliminary ideas into an outline before beginning the theme itself, to choose their words and illustrations carefully, and to develop good unity within paragraphs and transitions between them. They were also warned against using unqualified and fallacious statements. These qualities—as well as errors in spelling, punctuation, and sentence structure—were marked on the papers, and a few sentences of general evaluation were written at the end, including mention of commendable qualities as well as suggestions for improvement. Each paper was given two grades, one for content and organization, another for general correctness and accuracy.

The papers of the students in the Revision group were returned at the beginning of a class period. The general strengths and weaknesses of the essays were pointed out at that time, and excerpts exemplifying certain good features were read to the class to elicit comments on how the effectiveness was achieved. Then the students were required to correct the errors indicated on their papers while the reader went from student to student, giving assistance where it was needed. No more than one 50-minute class period was devoted to these procedures, the average time having been 35 minutes for each assignment—a total of some nine hours during the year.

The investigator reported that all of the essays of the Writing and Revision groups "were read by a group of qualified readers appointed by the investigator and paid on an hourly basis out of funds allotted to this investigation by the Alberta Advisory Committee on Educational Research." More specifically, the readers were graduate students, each specializing in the field of the instructor with whom he was working. The papers in art were thus read by a graduate student in art, those in music by a graduate student in music. The investigator noted that "most" of the readers "had had teaching experience," but he did not indicate how much. Presumably, few of them had been English teachers. The readers also conducted the follow-up work (described in the previous paragraph) undertaken when papers were returned to the Revision students.

But the regular instructors assigned the topics, discussed the assignment with the students, and collected the papers when they were due. The instructors also returned the papers to the Writing students.

In order to maintain consistency from one class to another, a general meeting of all the instructors was held in September, before the program was begun, to discuss the plan of the experiment. The investigator reported that the instructors "were interested in the investigation and eager to cooperate" when it was assured that they would not have the paper grading added to their usual duties. Sheets outlining the procedures for each group were distributed to all of the instructors and readers. (Copies are included in the final report as Appendices A, B, and C.) Furthermore, the instructors frequently conferred with the investigator about the topics they planned to assign, representative examples of which follow:

"What Music Means to Me"
"My Observation of Science Teaching in the Elementary School"
"My Visit to the Alberta Legislature" (and my conclusions)
"My Opinion of the Film 'Near Home' "
"Handwriting Can Be Improved"

After considering whether or not the theme grades should count as part of their regular subject grades, the investigator visited each of the six classes before the first assignment was given, "described the experiment, stated the responsibility of each group, and asked for their cooperation." He advised the students in the Writing and Revision groups that the practice would probably improve their writing, but he explained that essay grades would not affect their grades in the subject areas. "This approach seemed to prove quite satisfactory," the investigator stated, "because the students throughout the year made every effort to complete all assignments on time in spite of a rather heavy subject-matter program."

To determine whether or not the three groups were equivalent for the purposes of the experiment, all of the students were given three tests prior to the beginning of classes in September, 1956: (1) the Cooperative English Test A, Mechanics of Expression, Form 2, (2) the Cooperative English Test B2, Effectiveness of Expression, Form Z, and (3) an original essay written during a 50-minute period. Parallel forms of these three examinations were used again in March, 1957, to measure the performance of the students after instruction. In March, Form X of each of the Cooperative tests was used. These multiple-choice tests were administered on each occasion by the University of Alberta Student Advisory

Services, and the answer sheets were scored by IBM machine. The scaled scores were used in all computations.

The instructions for the essay pretest and post-test were given on mimeographed sheets (included in the final report as Appendices E and F). Because of the widely varying backgrounds of the students, a broad topic was assigned and suggestions were made on the instruction sheet of different ways in which the topic could be narrowed. The topic of the first of these 50-minute themes was "High Schools," and the topic for the post-test was "My Opinion," the latter permitting practically any kind of stand on any subject. On the instruction sheet for the post-test, the students were advised "Since your name will be replaced by a number before your essay is read, all opinions expressed are entirely anonymous so far as your readers are concerned."

Reliability in rating the essays was sought through a number of precautions. The anonymity of the student writer was preserved by having a colleague mix the 257 examination booklets so thoroughly that "both class and alphabetical arrangement were eliminated." Each paper was then numbered (from 1-257) on the cover and the first inside page, and the cover (with the student's name and number) was removed.

The essay tests were graded by two raters. One was the investigator; the other was the Chairman of the Grade Twelve Essay Marking Committee of the Alberta Department of Education. Both raters had had more than twelve years' experience in teaching English in high school. A score sheet was prepared to guide the essay test raters. The sheet provided for reading each theme according to general areas of consideration, under each of which its constituent factors were listed and the maximum number of points to be awarded each factor was given. A feature which may have contributed strongly to the validity of the experiment is that the score sheet went through several preliminary stages as the raters practiced with it on trial themes. (The fact that the raters helped develop the score sheet as they practiced with it suggests that they not only understood it but believed in it and followed its instructions.) The final score sheet (reproduced as Figure 1) contained fifteen categories, each of which could receive a maximum number of points.

In addition, each paper received a "basic mark" of 75 points, and deductions amounting to as much altogether as 100 points could be made for spelling, punctuation, usage, grammar, sentence, and form. The maximum total theoretically possible was 300 points.

Figure 1

Score Sheet to Be Used by Rating Committee

A. MARKS AWARDED:	Max.	Student
1. **Basic Mark (Give every paper):**	75	75
2. **Material:**		
a. Significance:	30	
b. Evidence of Critical Thinking (defining terms; recognizing generalizations; providing evidence, etc.):	30	
c. Originality (use of humor; exaggeration for effect; mock seriousness; anticlimax; understatement; pretentious language used for effect; etc.):	30	
3. **Organization:**		
a. Title (interest and appropriateness):	5	
b. Introduction:	10	
c. Logical sequence of paragraphs:	10	
d. Unity within paragraphs:	10	
e. Transition within paragraphs:	10	
f. General coherence:	10	
g. Effective conclusion:	10	
4. **Sentences:**		
a. Variety in sentence structure:	20	
b. General fluency:	10	
5. **Diction:**		
a. Exactness and vividness of nouns, verbs, adjectives, etc.:	20	
b. Interesting and appropriate figures of speech, comparisons, illustrations, allusions, quotations:	20	
	300	

B. DEDUCTIONS:

Spelling	1111111111111111111111111	
Punctuation	111111111111111	
Usage	111111111111111	
Grammar	111111111111111	
Sentence	111111111111111	
Form	111111111111111	
	TOTAL:	

C. ERROR RATE:

Length: ______________ words. Number of errors: ______________

Error Rate ______________ x 1,000 ______________

Practice rating was provided the two raters. Themes were obtained for practice by having fifty freshmen not in the experiment write on the same topic and under the same conditions as the students in the experiment. The two raters marked and discussed each of these papers, revised the score sheet, and then began rating the experimental papers. (The report does not make clear whether or not the raters were aware which papers were pretests and which post-tests. Since the papers were numbered from 1 to 257, we may infer that the pretests and post-tests were rated separately and hence that the raters knew which was which, though obviously they did not know which papers were written by the Control group, which by the Writing, and which by the Revision. Consequently, the results of the rating may be considered useful for comparing the relative composition quality of the three groups but not for measuring progress from pretest to post-test.)

Each rater read each paper at least three times:

> During the first reading, the errors in spelling, usage, punctuation, and sentence structure were noted, and an attempt was made to evaluate the content, introduction, and conclusion. During the second reading, examples of effective diction, concreteness, figures of speech, and critical thinking were recorded on the back of the scoring sheet and marks given for these. During the third reading, an attempt was made to evaluate the logical sequence of paragraphs, transition between paragraphs, unity, and the general coherence of the whole. Frequently a fourth reading was necessary to assess the accuracy of scores given.

The two raters independently marked and graded each paper, leaving to "a committee" the computation and recording of final scores. The raters jointly reviewed every twenty-fifth paper, as they went along, to insure continuous consistency but not to change the grade. Interpretations which had to be made of the score sheet as the rating progressed were reported (on pages 39-41). As a check on accuracy, each rater marked mechanical and grammatical errors in the margins, but "the number of errors was not considered in the awarding of marks for positive qualities." In order to adjust for the number of opportunities for error in longer and shorter themes, the number of errors was divided by the number of words written (and then multiplied by 1,000) to find an "error rate." The final score for each rater was computed by adding the 75-point "basic mark" and the points awarded for material, organization, sentences, and diction; then, from that possible maximum of 300 points was subtracted the "error rate." The final scores of the two raters were averaged and then divided by 3 to yield a *scaled score*—the number of points based on 100, which facilitated comparisons of the essay and objective test scores and made other statistical computations more manageable. The raw scores, subsequent adjustments, and resulting scaled scores for each of the 257 papers are given in Appendix J, though the scores for two post-test themes were estimated by the missing plots technique.

Rater reliability coefficients estimated by the Pearson product-moment formula were .91 for the pretest themes and .88 for the post-test themes, indicating a high degree of consistency between the two raters.

Results

The mean scores for each group on each of the three tests are given in Table 2. (The rest of this discussion will concern only the essay examination, not the two objective tests, other than to note here that the differences among the three groups on the two objective pretests were found not to be significant at the .05 level of confidence.)

Table 2

Mean Scores of Pretests and Post-tests on All Tests for All Three Groups

	Control Group (N=86)		Writing Group (N=86)		Revision Group (N=85)	
	Pre	Post	Pre	Post	Pre	Post
Mechanics of Expression	53.29	56.99	53.78	59.88	52.89	58.18
Effectiveness of Expression	57.05	58.05	57.16	59.09	43.35	57.28
Essay	42.65	48.28	42.10	48.97	40.67	51.79

Before analyses of variance and covariance could be conducted with the essay test scores, it was necessary to determine the degree of homogeneity of the variances among the three groups—Control, Writing, and Revision. Then the significance of the differences between the post-test means was calculated by an analysis of covariance to adjust the post-test means in terms of differences in the pretest means. Significance was determined by applying the t-test to the differences between the adjusted post-test means.

When the investigator computed F-ratios as tests for the homogeneity of the within-groups variances on the essay pretest, the investigator found one ratio significant at the .05 level—the ratio of the Control group to the Revision group. Consequently, he applied Bartlett's test of the significance of the differences among the within-groups variances. The value of the chi-square resulting from the calculation was 2.818, when it would have had to be 5.991 or greater to be significant for two degrees of freedom. This indicated that an analysis of variance and covariance was justified.

Analysis of variance on the essay pretest yielded an F-ratio of .8668, where 3.035 is significant at the .05 level of confidence with 2/252 degrees of freedom (and 4.695 is significant at the .01 level), indicating that there was no significant difference between any two groups at the beginning of the experiment. Analysis of covariance of the essay post-test scores, adjusted for differences in the pretest scores, yielded an F-ratio of 4.97 for the among means mean squares divided by the within-groups mean squares where 4.695 is significant at the .01 level of confidence with 2/251 degrees of freedom. Inasmuch as the three adjusted post-test means were thus seen to differ significantly, the investigator applied the t-test to the differences between pairs of adjusted post-test means. The difference between the means of the Writing and Control groups was significant at the .01 level, and the difference between the means of the Revision and Writing groups was significant at the .05 level of confidence. In other words, the Revision group improved very significantly over the Control group and significantly over the Writing group, though the Writing group did not improve significantly over the Control group.

The investigator explored further to determine whether or not the differences in achievement on the post-tests were in fact the result of the different experimental methods of instruction. He obtained the mean scores achieved by all the students on their final examinations in all their subjects:

Control Group	63.3
Writing Group	64.4
Revision Group	63.2

After determining by Garrett's method that the within-groups variances were homogeneous, the investigator computed the significance of the differences between means. Where t = 1.99 at the .05 level of confidence for 85 degrees of freedom, the critical ratio between the Writing and Control groups was 0.18, between the Revision and Control groups was 0.13, and between the Revision and Writing groups was 0.18. In other words, it is not apparent that instructional factors other than those in the experiment affected the results.

The investigator conducted one last analysis—an analysis of the gain in mean raw score from pretest to post-test essay made by each group on each of the fifteen categories on the score sheet. The gain was computed for each student in each of the fifteen categories. The mean gains and standard deviations are given in Table 3.

Table 3

Mean and Standard Deviation of Gain-Scores by Groups and Categories

Category	Control Group		Writing Group		Revision Group	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Significance	5.07	8.22	3.14	8.13	6.68	8.70
2. Critical Thinking	1.52	6.64	.58	6.09	1.04	8.64
3. Originality	1.16	6.66	1.23	6.57	2.24	8.34
4. Title	1.60	4.20	2.80	4.02	5.20	4.30
5. Introduction	.46	3.92	1.12	3.90	2.81	4.16
6. Logical Sequence	1.20	5.80	1.76	3.26	2.26	4.54
7. Unity	1.12	3.52	1.24	2.96	1.68	3.76
8. Transition	1.89	6.22	1.76	6.60	1.66	6.54
9. Coherence	.96	2.58	.46	2.88	1.26	3.24
10. Conclusion	1.08	5.12	1.94	4.92	2.12	5.76
11. Variety in Sentence Structure	1.97	7.11	2.99	7.08	4.28	6.30
12. General Fluency	.64	2.58	1.20	2.98	1.46	2.26
13. Diction	2.03	10.86	4.34	10.35	5.93	10.56
14. Figures of Speech	4.16	9.15	5.93	8.31	8.46	10.86
15. Error Rate	5.32	14.40	7.57	12.36	9.13	14.04

From the figures in Table 3, the investigator estimated the significance of the difference between each mean difference and the smallest mean difference (.46), using the usual critical ratio method. Using the 5 percent level of confidence, the investigator found those critical ratios significant which are shown in Table 4.

The Revision group gained significantly over the Writing group in three categories and over the Control group in six of the fifteen categories. (As has been shown on page 65, because the raters evidently knew which papers were pretests and which were post-tests but not which were written by students of the three respective groups, these "gains" should be viewed as evidence of the relative effectiveness of the three

Table 4

Significant Critical Ratios for Groups and Categories

Category	Control	Writing	Revision
1. Significance of Material	4.95	2.88	6.39**
2. Critical Thinking	....	....	....
3. Originality	....	....	....
4. Title	2.07	4.41	8.78***
5. Introduction	....	....	4.35***
6. Logical Sequence of Paragraphs	....	2.77	5.12
7. Unity	....	....	2.39
8. Transition	....	....	....
9. Coherence	....	....	....
10. Conclusion	....	2.43	2.41
11. Variety	....	3.09	5.09*
12. Fluency	....	....	2.56*
13. Diction	....	3.34	4.67*
14. Figures of Speech	3.59	5.88	6.56*
15. Error Rate	3.09	5.27	5.76
Total:	4	8	11

*Gain of Revision over Co[illegible] significant
**Gain of Revision over Wri[illegible] significant
***Gain of Revision over both Control and Writing significant

methods but probably not as clear evidence of absolute gain in a particular category.)

Among the conclusions which may be clearly drawn from this experiment are these:

1. Raters who use a scoring system which they understand and endorse, who have practiced rating papers by this system, and who make periodic checks on their rating may achieve a high degree of consistency in rating essays, or themes.
2. Since the students in the three groups did not differ significantly in their gain on the two objective post-tests but did differ significantly in their gain on the post-test essay examination, it may be seen that essay examinations can be used to measure changes which may not be measured by objective tests.
3. College freshmen whose writing is graded and thoroughly marked and criticized and who revise their papers in the light of

these matters can improve their writing more than college freshmen whose writing receives a few general suggestions but no grades or intensive marking and who do not revise their papers. (It is not clear, however, what the relative influence is of each of these three factors.) Further experimentation would be necessary, of course, before one could feel confident in making such a generalization about, say, junior high school pupils.

The Harris Study

Harris, Roland J. "An Experimental Inquiry into the Functions and Value of Formal Grammar in the Teaching of English, with Special Reference to the Teaching of Correct Written English to Children Aged Twelve to Fourteen." Unpublished Ph.D. dissertation, University of London, 1962. 291p.

Problem

This study investigated the relative teaching usefulness of what might be loosely referred to in the United States as "formal grammar" and a "direct method" of instruction. Although previous investigators have focused on this problem, most of them have either restricted their data to objective test results or have conducted their studies over too short a period of time. Harris' research was based on frequency counts of actual writing done before and after a two-year period of instruction. The reader's attention is also directed to the manner in which Harris tried out his procedures in a three-month pilot experiment and validated his criteria of measurement before he undertook his regular experiment.

The investigation compared two forms (or classes) of pupils in each of five London schools; ". . . one form had as nearly as possible the same general English work as the other, but with the one major difference that while one form studied and applied the terms of formal grammar in its composition work, the other used none of these terms and devoted the time saved to direct practice in writing." Although the investigator refers to the two groups of pupils as the "Grammar" and the "Non-Grammar" groups, actually grammar and composition were taught in each group. To prevent misunderstanding by American readers, the terms "Formal Grammar" and "Direct Method" will be used here to distinguish between the two groups. The essential differences were, as the investigator explained on pages 131-139, that the Formal Grammar group followed a logically organized program of traditional grammar instruction "through the parts of speech, with stress on the function of words"[1] and employed the traditional grammatical terminology in classroom teaching and in correcting compositions, and the Direct Method group evidently used no textbook or grammatical terminology but considered the elements of "sentence building and structure" which came to the teachers' attention as they read the children's writing, treating common errors in the classroom and in compositions "by means of example and imitation, instead of by the abstraction and generalization of the approach through formal grammar–which did not itself, of course, exclude the use of examples." The time saved by not teaching formal grammar in the Direct Method group was devoted to additional composition work.

To insure that the distinction between the two groups, and hence the problem of the experiment, is clear, an extended quotation from the final report is included here:

> Such mistakes as "Jim and me was going into the cave" . . . would be tackled in the grammar forms by direction to agreement of subject and verb and to the proper use of the nominative forms of pronouns when the pronouns are used as the subject. First, of course, subject, verb, agreement, pronoun, nominative or subject case, were terms that had to have been absorbed—at least, this was the ideal supposition. Probably the example which followed ("Jim and I were going. . . .") actually drove home the point, and one hoped that the rule would spread the influence of the example to other parallels. In the non-grammar forms, there was time for more examples. The sentence would be looked at in this way:
>
> "Jim and me was going into the cave."
>
> *Teacher:* "Would you say 'We was going into the cave'?"
> General dissent—one or two brave individuals aver that they would or might, and why not sir? But the vast majority of the class bring home to them the weight of convention. They admit to being wrong, which was more than was asked of them.
> *Teacher:* "What would you say then?"
> *Class:* "We were going into the cave."
> *Teacher:* "How many is 'we'?"
> *Class:* "Depends upon how many there are. More than one, anyway."
> *Teacher:* "Well, 'Jim and me' means more than one. So they must be followed by 'were.' Let's try some more examples. . . ." And the class, and the teacher, provide twenty or thirty examples in five minutes orally. With a group that needed this sort of practice, it was unwise to try to correct the pronoun error in the same lesson. One has to decide on priorities.

---

[1] The textbook used in four of the schools was Humphreys and Roberts' *Active English Course, Books 1 and 2* (London: University of London Press, 1939), and the text used in the remaining school was Allen and Mason's *An English Grammar of Function, Book 1* (London: Edward Arnold and Company, 1939).

Procedures

To refine his procedures and measuring instruments, the investigator conducted a three-months' pilot experiment with a pair of classes, first-year forms at a Technical School. The classes contained 22 and 24 boys. As in the main experiment, the plan for each class was the same except that the one lesson each week devoted to formal grammar in the one class was spent on writing and functional grammar in the other. The work done by both groups included "practice in summarising, in comprehension, in reading and composing verse, in composition, in silent reading, reading aloud and discussion." The Formal Grammar group studied sentence-types, subjects and predicate, the parts of speech, and the idea of function, using *Active English Course, Book 1,* the same text used by four of the five schools in the main experiment. Instead of studying those matters, the Direct Method group wrote a story about the exploits of a boy separated from his parents during the war. Evidently the class worked out a common plan for the story but each pupil wrote his own version, doing a chapter each week. The teacher afforded instruction in planning the story, achieving inherent plausibility (even if not factual accuracy), maintaining suspense, and using details to suggest character and establish atmosphere. No class or group instruction was given the Direct Method group in the imitation of sentence structures; but individuals were aided in rephrasing their sentences, and the results were "often brought to the attention of the class." When the pupils finally completed their stories, they exchanged them and evaluated them in committees of three, for suspense, characterization, and vividness of scenes. The investigator noted that the procedures used with the Direct Method group took as much but no more time than those used with the Formal Grammar group and that they elicited "not only enthusiasm but also a certain self-criticism and purposive modification of habits of writing."

Two measures were administered to each group before and after the three-months' instruction. One was a short-answer formal grammar test, also used in the main experiment and included in the report as Appendix 3, which required the identifying in sentences and naming from recall of the parts of speech, subject and direct object, phrase and clause, tense and mood, and the like. The examinees were also required to explain *"in grammatical terms"* what was wrong with a number of sentences. The Formal Grammar group showed much better progress than the

Direct Method group on this measure: the difference between means was 7.345, the standard error was 1.349, and t was 5.45.

The other measure was a composition. Both classes wrote a descriptive paper on "A Day in the Country" at the beginning and on "A Day at the Seaside" at the end of the three-month period of instruction, being afforded two 40-minute class periods for the writing each time. Instead of evaluating the papers by solely subjective judgment, the investigator conducted a number of frequency counts, using from each paper the first 500 words to the nearest end of a sentence. Two types of counts were made: (1) the number of words per error, as listed in Table 1, and (2) the number of complex or noncomplex sentences and the number of correct or incorrect sentences. Illustrations were not given of which types of usages and sentences were considered correct and which incorrect,

Table 1

(Appendix 1 in the Harris Report)
A List of Common Errors in Children's Writing

1. Certain errors in punctuation, notably—
   a. omission of question mark.
   b. omission of full stop.
   c. omission of comma between items in a list, words in apposition, main and non-defining clauses.
   d. omission of apostrophe in 's for possession.
2. Failure to use capital letters—
   a. at beginning of sentence.
   b. in proper nouns and adjectives.
3. Misuse of various parts of speech—
   a. adjective or preposition as adverb.
   b. wrong comparatives and superlatives.
   c. faulty positioning of adverbs.
   d. failure to give pronouns a clear antecedent.
   e. wrong or ambiguous use of prepositions.
   f. mismanagement of conjunctions.
   g. misuse of object forms of pronouns.
   h. failure to use relative pronoun or adverb.
4. Failure to give a finite verb to each clause, error in or omission of any important verbal word.
5. Lack of agreement between verb and subject.
6. Faulty sequence of tenses.
7. Unrelated or false participle, e.g., "laying" for "lying."

NOTE: In the error count referred to in the pilot experiment, and in the final reckoning of the third measuring instrument, errors 1a, 1d, 2a, and 2b were omitted from the calculations as being too remote from grammar as such and too simply mere matters of convention which training in formal grammar should not be expected to help in eradicating.

but a familiarity with examples elsewhere in the Harris report (especially in Tests A and B in Appendix 3) suggests that a rather strict sense of formality was observed. (For instance, a distinction in meaning is drawn on page 134 between "He only found three ducats" and "He found only three ducats.")

The differences between the means of the Formal Grammar and Direct Method groups on the frequency count measures were all small, insignificant, and favoring the Direct Method group. Encouraged by those results, considering the small number of children examined and the shortness of the pilot period of instruction, the investigator decided to use the frequency count approach in the main experiment but to refine the procedures involved. But he did not reach that decision until he had tried to gauge the "differences in correctness" between boys of 12 and 14 years of age by using two tests similar to those common in the United States, tests in which the examinees choose which of two alternate forms is the better. Finding no regularity of increase in scores with age when he administered the tests to boys aged 11, 12, 13, and 14 at one school, the investigator abandoned that type of test. The formal grammar test (Test C), on the other hand, not only achieved a coefficient of reliability of .923 ± .014 by the split-half method, but it yielded regular increases in scores with age, as Table 2 shows.

Table 2

Validation Data for *Test C*

Age Group	Average Score	Range of Scores
11 years	8.1	0-16
12 "	13.0	4-22
13 "	22.8	12-34
14 "	33.1	19-48

In order to find some frequency count measures which would reflect "the rate of growth of a child's maturing style" as well as Test C reflected maturity in knowledge of formal grammar, the investigator made a comparative analysis of some compositions written by 63 ten-year-old children and by 25 fifteen-year-olds. After a number of initial assessments which did not prove to differentiate satisfactorily between the two groups, the investigator developed a set of 11 critieria "which occurred sufficiently often to give a clear measurement, and which did not overlap too much." The 11 criteria are given in Table 3. For each criterion, the investigator had computed the difference between the mean scores

Table 3

Criteria of Maturing Style

a. Average length of the correct simple sentence; i.e., the simple sentence containing none of the errors listed in Table 1. (D 4.86, SE 0.832, t 5.84)
b. Instances of the omission of the full stop at the end of a sentence and of inadequate separation between clauses, scored by subtracting the number of omissions from the number of sentences in which the stop was correctly inserted. (D 11.41, SE 1.670, t 6.79)
c. Number of words per common error. (See Table 1, including the note at the foot.) (D 141.0, SE 22.90, t 6.16)
d. Number of different sentence patterns used by each writer, counting only correct sentences but accounting, for example, different types of adverbial clauses or inversion of the order of similar clauses as sufficient to provide a new pattern. (D 6.31, SE 1.12, t 5.66)
e. Number of non-simple sentences, whether correct or incorrect, from which the number of simple sentences was subtracted. (D 7.00, SE 1.21, t 5.78)
f. Number of subordinate clauses used—the actual number (not in proportion to the number of main clauses). (D 11.44, SE 1.71, t 6.67)
g. Total number of words written in a limited time. (D 235.4, SE 23.00, t 10.24)
h. Number of correct simple sentences, minus those containing an error (even though minus scores were frequent at 10 and 11 years of age). (D 9.513, SE 0.916, t 10.38)
i. Number of correct simple sentences containing two or more qualifying phrases, each phrase consisting of three or more words. (D 1.063, SE 0.251, t 4.24)
j. Total number of correct sentences minus the number of incorrect. (D 13.68, SE 1.373, t 9.96)
k. Number of adjectival clauses and phrases, each phrase consisting of three or more words. (D 12.97, SE 1.341, t 9.67)

at age 10 and age 15, the standard error, and the critical ratio (t). The three statistics are given in parentheses at the end of each item in Table 3, demonstrating the apparent significance of each of the items finally selected.

To discover how reliable were the differences established by the criteria, the investigator applied them to two descriptive compositions written, a week apart, by 27 thirteen-year-old boys who were allowed forty minutes to write on each of these topics: "A House on Fire" and "A Foggy Night." Table 4 shows the product moment correlation of their scores (r), the standard error, and the reliability quotient (t), with the level of significance. It is evident that all criteria except a and g reached a good or high level of reliability.

Other investigators interested in conducting frequency counts using Harris' 11 criteria may also wish to use the simplified process he developed for tabulating his counts and scores, a sample of which is given in Figure 1.

Table 4

Reliability of the Criteria

	r	SE	t	level of signif.
a. simple sentence length	+0.239	0.181	1.23	worse than 0.1
b. omission of full stop	+0.574	0.129	3.49	better than 0.01
c. words per common error	+0.844	0.055	7.87	better than 0.001
d. number of sentence patterns	+0.791	0.072	6.46	better than 0.001
e. non-simple sentences	+0.393	0.163	2.14	better than 0.05
f. number of subordinate clauses	+0.705	0.097	4.97	better than 0.001
g. total words	+0.232	0.178	1.21	worse than 0.1
h. complex sentences	+0.685	0.102	4.70	better than 0.001
i. simple sentences with two or more phrases	+0.447	0.154	2.49	better than 0.02
j. total correct sentences	+0.839	0.057	7.71	better than 0.001
k. adjectival clauses and phrases	+0.446	0.154	2.49	better than 0.02

With his measuring instruments prepared, the investigator completed the design of his experiment. After some difficulty from other schools because of the length of the experiment, the investigator secured the cooperation of five London schools representing a range of socioeconomic conditions: two from West London (Isleworth and Dormers Wells), one from nearer the center (Christopher Wren), and two from "rather East" (Shoreditch and Owens). See Table 5.

Because of administrative necessities at the five schools, the investigator was not able to select the pupils for each group and hence equate the groups for intelligence, background, or success in English. Fortunately, however, the pupils had not been grouped according to ability, and the pairs of classes in each school were roughly equivalent (See Table 34, page 278, in the original study), with the Formal Grammar groups having in three of the five schools a slight superiority in IQ. Although some pupils were lost from the experiment during the second year, 119 were able to spend both years in the Formal Grammar classes, 109 in the Direct Method.

Figure 1

Paper No.	Sentence No.	Error No.	Clauses Used	No. of Words	S—	S+2	Adj. Phrases
2	1		M Vt Vt Nobj	29			
	2		Vt M M Vt Vreas	31			
	3		Vt M J J	20			
	4		M Nobj M Vt	16			
	5	3f,3f	Vt M Nobj M M M M	34			
	6		Vt M	10			
	7	1b	M M M Vreas M Nobj	31			
	8	1b	M M M	17			
	9		M M M M	22			
	10	1b	M M M M Nobj M	40			1
	11	6,3d	M M Vt Vt Nobj	34			1
	12	6	M M M M	24			

Criterion Scores:	a	b	c	d	e	f	g	h	i	j	k
	—	+6	38.5	6	12	20	308	+1	—	0	4

Key

S— = simple sentence without any qualifying phrase.
S+2 = simple sentence but with two or more qualifying phrases.
M = main clause.
N = noun clause.
Nobj = noun clause as object.
Nappos = noun clause in apposition.
V = adverbial clause.
Vreas = adverbial clause of reason.
Vconc = adverbial clause of concession.
Vt = adverbial clause of time.
J = adjectival clause.

Table 5

Participants in the Experiment

	Formal Grammar Classes	Direct Method Classes
Grammar Schools		
Isleworth	29 boys	26 boys
Owens	21 boys	25 boys
Secondary Modern		
Dormers Wells	29 girls	21 girls
Comprehensive/Technical		
Christopher Wren	21 boys	23 boys
Shoreditch	19 boys	14 boys
	119	109

In four of the schools, one teacher taught both classes, but in the Secondary Modern school the two classes were taught by different individuals. Though the investigator acknowledged that "no certainty exists that each teacher was absolutely impartial in his attitude to the teaching or abandonment of formal grammar, or indeed equally adept at the one approach as at the other," he judged that "all the teachers were willing to be convinced by evidence, desirous of finding out the truth about the matter in hand, competent and practised in the present grammar of the language." (No further attempt seems to have been made to estimate the probable effects of the teacher attitudes and skills on the outcome of the experiment.) The investigator also reported "The general method of grammar teaching among the five teachers was surprisingly uniform"—grammatical form and terminology being "closely tied to function and meaning," but he did not describe any procedures used to observe or control the teaching in either group, none, that is, other than the teacher direction for the Formal Grammar groups implied by the textbooks, the formal grammar test, the List of Common Errors, and "the relevant sections of the Course of Study." The teacher instructions for the Direct Method groups "excluded any reference to formal grammar and specified attention to practice in writing and imitating conventional forms and structures." The investigator also stated that the teachers were careful not to use the terminology of formal grammar when correcting the homework of the Direct Method pupils.

Each class met for five 40-minute periods per week and divided its work during four of the periods among the usual areas of reading, drama, poetry, and composition. The difference came in the fifth period, when one group emphasized formal grammar and the other focused on direct methods of instruction. (Details are given early in this summary, under "Problem.") The investigator described the difference further: "The work attempted in the [Direct Method] lesson was basically an extension of the usual composition practice [in Formal Grammar classes], but most teachers gave cohesion and interest to the work by engaging in a variety of longer projects such as the compilation of a diary, a form newspaper, an adventure story similar in general outline to the one discussed in the pilot experiment, or a book of hobbies." Thus some of the time released by omitting study of the terminology of formal grammar could be devoted to writing activity projects and to drawing illustrative sentences, points of usage, and paragraphs from the stories to teach the improvement of writing. A piece of continuous writing was also attempted in the Formal Grammar classes, but not much time was available for the project inasmuch as these classes were carefully following the integrated grammar and composition lessons in their textbooks.

Results

The investigator reported the results of the pretest and post-test in formal grammar (Test C) as shown in Table 6. He wrote, "It is evident from these scores that good progress was made by the grammar forms [Formal Grammar classes] in learning not only the terms of formal grammar, but how to apply these, for Test C contained questions requiring both recognition and use of the common grammatical terminology at the appropriate level." The gains of the Formal Grammar group were, of course, significantly higher on this test than the gains of the Direct Method group.

Table 6

Scores on Test C, the Formal Grammar Test

	Group	First Average	Second Average	Gain	Difference	SE	t
School A	Formal	33.03	49.71	16.68	16.11	1.955	8.24
(Grammar)	Direct	34.12	34.69	0.57			
School B	Formal	36.33	66.62	30.29	26.97	2.393	11.27
(Grammar)	Direct	34.88	38.20	3.32			
School C	Formal	20.10	34.34	14.24	12.86	1.825	7.04
(Sec. Mod.)	Direct	22.09	23.47	1.38			
School D	Formal	19.48	39.96	20.48	19.74	1.359	14.53
(Comp. Tech.)	Direct	21.09	21.83	0.74			
School E	Formal	15.84	32.84	17.00	15.50	1.669	9.29
(Comp. Tech.)	Direct	22.71	24.21	1.50			

Frequency counts for the criteria listed in Table 3 were made from essays written at the beginning and end of the two-year experiment. A similar analysis was conducted of essays written at three of the schools after nine months of instruction, "to check that things were proceeding according to plan." Although each class used the same topic for the final essay that it used for the first, other topics (but ones calling for the same mode of discourse) were employed for the intermediate composition. (The investigator did not report the topics used, but he indicated that the mode of discourse was narrative or descriptive in each case.)

It seems highly important, in the light of the results of the final composition (reported later), to note that the results of the intermediate composition analysis were very inconclusive. The differences between means of the Formal Grammar and Direct Method groups at each school were computed for each of the eleven criteria, a total of 33 differences. Of these 33, only four differences were significant, where t = 3 or over. Three of those favored the Direct Method group, one the Formal Grammar group. Another four differences, reaching a fair level of significance (t = 2.70 or over), all favored the Direct Method group. Of the total of 33 differences, 16 favored the Formal Grammar group and 17 the Direct Method group, leading the investigator to conclude "The period of nine months or one academic year is too short to expect any stable changes to have been made, and to allow for variation of pace within a course."

Significant differences appeared much more clearly in the comparative analysis of the compositions written at the end of the two-year experiment, as is shown in Table 7. Of the 55 differences computed then, 11 were significant, and all 11 favored the Direct Method group. Of the 25 differences represented by the five very reliable measures (c,d,f,h, and j), only six favored the Formal Grammar group, and none of those differences was significant.

Table 7

Final Scores on Frequency Count Criteria

Key

The level of reliability of each measure is indicated at the left by "V" (very reliable), "F" (fairly reliable), or "U" (unreliable). Where "g" is given, the difference favors a Formal Grammar class; unmarked t-scores indicate a difference in favor of a Direct Method class.

An asterisk (*) indicates that the difference reaches a significant level.

Rel.	Measure and School	Diff. bet. Means	SE	t	
U	a. Ave. length of simple sentences				
	A	1.717	1.296	1.32	
	B	0.975	1.212	0.79	
	C	3.190	1.944	1.64	g
	D	1.779	2.065	0.86	
	E	0.223	2.262	0.10	
F	b. Omission of full stop				
	A	2.171	1.873	1.16	
	B	3.345	3.180	1.05	
	C	1.766	1.712	1.03	g
	D	3.168	2.316	1.37	g
	E	0.046	3.065	0.02	g

(Table 7 cont'd)

Rel.	Measure and School	Diff. bet. Means	SE	t		
V	c. Words per common error					
	A	14.26	13.63	1.05	g	
	B	85.21	26.68	3.19		•
	C	44.41	13.32	3.33		•
	D	38.41	11.92	3.22		•
	E	41.46	21.84	1.89		
V	d. Different sentence patterns					
	A	0.322	0.685	0.47	g	
	B	3.331	1.042	3.19		•
	C	2.778	0.767	3.62		•
	D	2.340	0.900	2.59		
	E	2.824	1.149	2.46		
F-U	e. Non-simple sentences minus simple sentences					
	A	2.120	1.140	1.86		
	B	3.640	2.056	1.77		
	C	0.710	1.688	0.42		
	D	1.374	1.848	0.74	g	
	E	1.105	1.468	0.69	g	
V	f. Subordinate clauses					
	A	1.174	1.637	0.72		
	B	1.116	2.233	0.49		
	C	1.951	2.274	0.86		
	D	0.036	1.919	0.02	g	
	E	0.255	2.376	0.11	g	
U	g. Total words					
	A	70.43	21.54	3.27		•
	B	47.99	36.63	1.31		
	C	35.16	31.39	1.12		
	D	37.02	19.18	1.93	g	
	E	30.16	37.23	0.81	g	
V	h. Complex sentences					
	A	0.648	1.107	0.59	g	
	B	5.042	1.466	3.44		•
	C	4.495	1.185	3.08		•
	D	4.363	1.193	3.66		•
	E	6.327	1.779	3.56		•
F-U	i. Simple sentences with two or more phrases					
	A	0.076	1.899	0.04	g	
	B	0.436	2.493	0.17		
	C	0.889	0.393	2.26	g	
	D	0.417	0.473	0.89		
	E	0.918	0.793	1.16		

(Table 7 cont'd)

Rel.	Measure and School	Diff. bet. Means	SE	t	
V	j. Number of correct sentences				
	A	2.607	2.168	1.20	g
	B	10.230	4.225	2.42	
	C	1.807	2.581	0.70	
	D	8.276	2.451	3.38	
	E	8.308	3.434	2.42	
F-U	k. Adjectival clauses and phrases				
	A	3.454	1.322	2.61	
	B	2.791	2.408	1.16	
	C	2.003	1.820	1.10	g
	D	4.217	1.861	2.27	g
	E	2.850	2.767	1.03	g

In addition to the above computations, the investigator tabulated the number of papers containing each of the Common Errors listed in Table 1 and the total number of each of such errors in all papers in each group. Because the total number of each error has not been adjusted according to the total number of words, only the number of essays containing each error is given in Table 8.

From these results the investigator concluded, among other things, that there was "the lack of effective tie between a relatively high

Table 8

Number of Essays Containing Each Common Error

	First Essays		Final Essays	
	Formal	Direct	Formal	Direct
1b. Omission of full stop	107	87	91	76
1c. Faults in use of comma	96	84	103	74
3a. Adj. or prep. as adverb	1	2	7	5
3b. Wrong compar. or super.	1	1	2	1
3c. Faulty position of adverb	6	8	13	4
3d. No clear antecedent for pronoun	38	42	54	44
3e. Misuse of prepositions	21	15	48	19
3f. Misuse of conjunctions	53	44	63	39
3g. Misuse of obj. form of pronoun	7	8	15	6
3h. Failure to use relative	0	0	0	0
4. No finite verb or omission of an important verbal	37	37	50	21
5. Lack of subj.-verb agreement	18	14	28	22
6. Faulty sequence of tenses	0	2	16	12
7. Unrelated participles	30	27	36	18

grammatical score and improvement in the measured items of the essay"—the highest correlation between the top third of the scores on the final grammar test with changes in score between the first and final essays was merely .229 ± .138. The investigator also expressed dissatisfaction with the low levels of achievement of the Formal Grammar group on Test C, only one class having more than 50 percent of the answers correct. That low achievement, he wrote, "may be seen as a major factor throwing doubt on the advisability of studying formal grammar in the early part of the Secondary School." Noting that his students were drawn from a range of schools and socioeconomic backgrounds, he also concluded that "The failure to profit from instruction in [traditional formal] grammar is thus not confined to any one educational environment or category of children" (of those he studied). In short, the investigator concluded, "It seems safe to infer that the study of English grammatical terminology had a negligible or even a relatively harmful effect upon the correctness of children's writing in the early part of the five Secondary Schools." (Based as it was on the use of traditional grammar, the Harris study does not necessarily prove, of course, the ineffectiveness of instruction based on structural or generative grammar.)

The Kincaid Study

Kincaid, Gerald L. "Some Factors Affecting Variations in the Quality of Students' Writing." Unpublished Ed.D. dissertation, (Michigan State College) Michigan State University, 1953. 116p. University Microfilm No. 5922.

Problem

Noting the widespread practice of judging a student's writing by one sample of his work, the investigator sought "to determine whether a *single paper* written by a student on a *given topic* at a *particular time*" constituted a valid basis for evaluation. The investigation was based on the evaluation of 320 papers written between the fall term, 1951, and the winter term, 1952, by students enrolled in the first quarter of the Written and Spoken English course (later renamed Communication Skills) at Michigan State College (now Michigan State University). The reader's attention is directed especially to the superior design of this experiment and to the care with which the theme rating was planned.

Procedure

The students whose papers were evaluated in this investigation came from the classes of two instructors, who collaborated in the selection of topics similar in nature to writing assignments used previously in the term. In order to test the hypothesis that "any given topic provides the same stimulus as any other topic," some students wrote the same day on two similar topics, some on two dissimilar topics. The similar assignments used were "It Is (Is Not) Too Far Between Classrooms for Students to Travel During the Ten-Minute Period Allowed" and "Textbooks for Freshmen at Michigan State College Are (Are Not) Too Expensive," hereinafter referred to as "Distance Between Classes" and "Cost of Textbooks." Each student writing on one of these topics was "provided a brochure of information" pertaining to that topic. A third, dissimilar, assignment was also used: "Give Directions to a Stranger at the Union Building Enabling Him to Get to Building A-6 on South Campus," henceforth referred to as "Giving Directions." Each student writing on this topic was provided with a map of the campus. (Although these topics were selected to prevent content factors from affecting the quality of the writing, the nature of these assignments may lead some readers to question whether the topics were pitched at a level representative of assignments commonly used in other colleges and whether the material provided would tend to make the writing of various students more alike than it normally would be.)

A second hypothesis was "that any given topic elicits constant responses at different times." To test this, some students wrote on the same topic on different days, seeking to determine "whether variations in efficiency from day to day have any effect on the quality of student writing."

That "the psychological pressure of the examination situation has no adverse effect on the quality of student writing" was a third hypothesis. This was tested by having some students write without such a pressure on one day and with it, on the same topic, on another day. Paralleling them, a control group wrote on the same topic without any examination pressure on either occasion.

A fourth hypothesis was "that the quality of student writing is stable from topic to topic and from time to time, with or without the pressure of an examination, *regardless of variations in writing ability.*" The investigator tested this assumption by analyzing the variations in the quality of the writing in relation to the rank of the student's writing ability.

The design of the experiment provided for the variables and controls in the manner shown in Table 1.

Table 1

Design of Experiment

Key: Topic 1—"Distance Between Classes"
Topic 2—"Cost of Textbooks"
Topic 3—"Giving Directions"
W/O—Writing without the examination pressure
W —Writing with the examination pressure

	First Day		Second Day	
	Similar Topics		Similar Topics	
	Topic 1	Topic 2	Topic 1	Topic 2
Group A	W/O	W/O	W/O	W/O
Group C	W/O	W/O	W	W
	Dissimilar Topics		Dissimilar Topics	
	Topic 1	Topic 3	Topic 1	Topic 3
Group B	W/O	W/O	W/O	W/O
Group D	W/O	W/O	W	W

To prevent the intrusion of extraneous influences into the experiment, the investigator instituted a number of controls to allow for fatigue, motivation, and the like. Each of the four groups of students wrote all four papers during their regular two-hour laboratory periods, when they were accustomed to write. All four groups regularly met on the same days—Tuesdays and Thursdays. The presence of a minimum of twenty students in each group could confidently be expected. To eliminate the effects of fatigue, every other student wrote on the first topic during the first hour, took a ten-minute break, and then wrote on the second topic. The other students first wrote on the second topic and then, after a ten-minute break, on the first topic.

On the first day in which the students wrote, they were all instructed that their grades for the term would not be affected by the papers they wrote that day. As a stimulus to good writing, however, they were informed that grades would be assigned to their papers by a team of raters and reported to the students for comparison to their usual grades. On the second day of writing, the students in Groups A and B were given much the same instruction, again being assured that their grades would not count but in addition being informed that "The results of the grading will be posted on the bulletin board early next term." (If the students interpreted this to mean that their names would be posted with their grades, this procedure probably should be considered a substitution of social pressure for academic pressure, not an elimination of pressure.) The students in Groups C and D were informed in their instructions on the second day not only that their grades would be posted but that the grades would be considered as a part of their final examination and hence would affect their final grades for the term. It was also pointed out that, having written on the topics before, they "should be able to do a better job of writing" on them this time.

To prevent the papers written on the second day from being merely second editions of the first papers, the investigator did not wish to space them very close. On the other hand, he did not wish to space them so far apart that new learning experiences would cause more differences in quality of writing than variations in efficiency would cause. He favored the shorter time in his compromise, spacing the two days of writing one week apart.

To select reliable raters to evaluate the student themes, the investigator examined "the records of ratings by more than forty staff members of the Department of Written and Spoken English," selecting nine who "demonstrated a high degree of sensitiveness and reliability." Of these, three were "willing and able to devote the time necessary" to rate each of the 320 papers. All three raters judged all 320 papers, rating "at the same time and under the same external conditions," employing the usual rating scale for the Written and Spoken English course at Michigan State. Each paper was rated from 1 (unsatisfactory) to 10 (superior) on each of five equally weighted categories: conventions of grammar, sentence structure, diction, organization, and content. The average of the two closest ratings became the score of each paper. But if the three ratings were equidistant (22,24,26) or if no two ratings came within five points of each other, the average of all three ratings was taken as the score.

To preclude influences due to variations in rater efficiency from day to day, all three graders judged all four papers of each student during the same rating period; Rater 1 judged the first paper and passed it to Rater 2, who judged it and passed it to Rater 3. (The report did not state whether or not a grader could tell, at the time of rating, what score another grader had assigned the paper, but the investigator has advised the writers of this pamphlet that "No rater had any indication of a score assigned by a previous rater.") In addition, the papers were placed in such an order that any two themes by the same student were separated by the themes of at least two other students. (The investigator did not state whether or not the raters knew any or many of the student writers or could identify the papers of a particular group, but the investigator has advised that the students' names were removed from the themes and the papers were assigned numbers so that no grader knew what student's paper he was rating at any time.)

The raters judged papers for six hours each day, with two 90-minute periods each morning, separated by a 15- to 20-minute rest period, and a similar schedule each afternoon, for a total of three days. Some papers were eliminated from the study by random procedure as a means of equalizing the size of each group of students. These extra papers were studied for clues to the criteria by which the organization and content of the other themes on those topics could be evaluated.

As a means of seeing that the 80 students were representative of the freshman population at Michigan State, the investigator relied on the usual registration procedure for the course. Students had been admitted to registration in alphabetical order during the three-day registration period, and enrollment in each section was limited each day, automatically distributing throughout the alphabet the names of the students in each of the four groups.

It was at least as important for the four groups to be equal in their ability to write. The fact that all 80 students wrote their first theme on "Distance Between Classes" provided a basis for determining whether or not the four groups were equivalent. The mean scores of the first papers were as follows:

Group A	**Group B**	**Group C**	**Group D**
24.5	23.4	25.4	22.4

An analysis of variance produced an F-ratio of .7457 with 3/76 degrees of freedom where F=2.73 for significance at the 5 percent level of confidence and F=4.06 for significance at the 1 percent level of confidence. In other words, the indication was very strong that each of the four groups was representative of the total group of 80 students.

Results

To test whether or not fatigue did reduce writing efficiency, the investigator compared the scores on the Topic 1 papers written during the first hour to the scores on the Topic 1 papers written during the second hour. Since the alternate topics were distributed systematically to every other student, it was reasonable to assume that the two groups of students were equal in writing ability. The mean score of the Topic 1 papers written during the first hour was 23.735, during the second hour was 23.925, indicating that fatigue was not a factor or that it was slightly more than offset by another, unknown factor—perhaps a "practice effect." Application of the t-test to the difference of .19 indicated that such a difference would happen by chance alone between 40 and 50 times out of a hundred cases. The investigator then concluded that "neither the fatigue factor nor any other factor relative to the first or second hour of the writing period had any significant effect on the average quality of the students' writing."

To determine whether or not the day-to-day efficiency or the differences in topics were affecting the caliber of the students' writing, the investigator compared the mean scores of the Group A papers and the mean scores of the Group B papers, both of these groups having written on both days without examination pressure. The means for Group A, writing on similar topics, were these:

First Day		Second Day	
Topic 1	**Topic 2**	**Topic 1**	**Topic 2**
24.5	23.7	21.9	21.3

The F-ratio obtained for 3/76 degrees of freedom was .7480, where F=2.73 for significance at the 5 percent level and F=4.06 for significance at the 1 percent level. The means for Group B, writing on dissimilar topics, were these:

First Day		Second Day	
Topic 1	**Topic 3**	**Topic 1**	**Topic 3**
25.4	21.4	24.1	21.5

Here the F-ratio obtained for 3/76 degrees of freedom was .9837—not approaching significance at the 5 percent level. In short, there being no significant difference between the means of any two sets of papers written by Group A or any two sets by Group B, the investigator concluded that, with *groups* of writers, there were no significant effects on the quality of writing as a result of variations in day-to-day efficiency or variations in topics assigned. (The reader should be reminded here that the three topics were all simple topics for college freshmen and that standard information was supplied to the students. Obviously, different results might well have been obtained if topics had been used which demanded more complicated analysis or more specialized knowledge. Remembering that Topics 1 and 2 were argumentative in nature, Topic 3 expository, one might reasonably conclude that this *variation in mode of discourse*—rather than variation in topic—had no significant effect on the average quality of the writing of a group of students.)

As a test of the effect of examination pressure on the writing of groups of students, the investigator compared the mean scores of the Group C papers as well as the mean scores of the Group D papers. The primary variables were the topics (similar for Group C, dissimilar for Group D) and the psychological pressure of the examination situation (applied to both groups the second day). Again, analyses of variance were used to seek significant differences among the mean scores of the four sets of themes for each group. The means for Group C, writing on similar topics, were these:

First Day		Second Day	
Topic 1	Topic 2	Topic 1	Topic 2
23.4	23.9	25.0	24.6

The F-ratio with 3/76 degrees of freedom was .2700, not approaching significance at the 5 percent level. The means for Group D, writing on dissimilar topics, were these:

First Day		Second Day	
Topic 1	Topic 3	Topic 1	Topic 3
22.4	19.2	23.9	20.9

The F-ratio with 3/76 degrees of freedom was 1.48, not approaching the F=2.73 needed for significance at the 5 percent level. With no significant difference found among the means of the four sets of papers written by either Group C or Group D, the investigator concluded that neither the examination pressure, content factors (from the use of similar and dissimilar topics), nor a combination of such factors significantly affected the average quality of writing of these students.

In a further attempt to test the effect of psychological pressure, the investigator combined the scores of Groups A and B on Topic 1 on each day (writing "without pressure" either day) and combined the scores of Groups C and D on Topic 1 on each day (writing "with pressure" the second day). Reasoning that this combination of scores was justified by the lack of significant differences among the four groups on Topic 1 on the first day and by the doubling of the number of student papers which could be used in testing the effect of pressure, the investigator obtained these means:

First Day		Second Day	
Groups A and B	Groups C and D	Groups A and B	Groups C and D
24.9	22.9	22.5	24.5

An analysis of variance yielded an F-ratio of .7957 with 3/156 degrees of freedom where F=2.67 for significance at the 5 percent level of confidence. This analysis thus provided further evidence that examination pressure "had no effect on the average quality of writing by the groups of students involved" and, in the case of Groups A and B (writing without examination pressure on either day), that there was no significant variation in level of efficiency from day to day in the *average* quality of writing by these students.

The investigator pointed out that his preceding findings concerned *groups* of students but that individual variations may be hidden by cancelling each other in group means. Consequently, he also examined differences between scores on papers by the same student. Recognizing that such differences may be produced by the unreliability of ratings as well as by variations in the quality of student writing, and further recognizing that ratings by the same two raters were not always represented in the scores on papers, the investigator used a procedure developed by Robert Ebel for determining the reliability and error for sets of ratings.[*] First the investigator computed the overall reliability and standard error of the ratings "in order to make more accurate comparisons of a student's obtained scores on different topics":

	Standard Error	Reliability
Topic 1	3.2031	.80
Topic 2	3.4583	.77
Topic 3	2.5416	.91
Total	3.1203	.83

Then, using a procedure described in detail on pages 75-76 of the study, the investigator determined that differences between two papers greater than certain "theoretical differences" would have to be obtained for the difference in quality of writing to be significant:

---

[*] Robert L. Ebel, "Estimation of the Reliability of Ratings," *Psychometrika*, XVI (December, 1951), 407-424. "Ebel points out that if decisions are to be made in practice by comparing averages which come from different groups of raters, then the 'between-raters' variance should be included as a part of the error term." G.L.K.

	Theoretical Difference	Minimum Difference Used
Topic 1	8.88	9.00
Topic 2	9.58	10.00
Topic 3	7.04	7.50
Total	8.65	9.00

Each of the following pairings of the four papers by each student was examined to see whether the differences equaled or exceeded the pertinent minimum difference:

1. the first topic on different days
2. the second topic on different days
3. the first and second topics on the first day
4. the first and second topics on the second day

The difference for each student on each pairing is reported in Appendix C of the study. Since "no more than 16 such variations could be expected as a result of the unreliability of the ratings" (the standard error of the difference between paired scores having been computed at the 5 percent level of confidence), it is striking to note that 86 significant variations in quality occurred out of 320 opportunities (4 pairings times 80 students). The 86 variations were distributed among 47, or over 58 percent, of the students involved.

The 80 students were also ranked according to their writing ability, using the student's highest score (of the four papers) as a basis for the ranking, reasoning that the consequent distortion resulting from error in the ratings was not as serious in this study as distortion resulting from variations in performance. However, as a check against the first type of distortion, the investigator also reranked the students according to the average of all four scores. The reranking resulted in a shift of only four students from the top half to the bottom half of the group. However, while 15 of the 20 students lowest in the first ranking remained in the bottom fourth when reranked according to average score, only 11 of the 20 students highest in the first ranking remained in the top fourth when reranked, suggesting that quality of writing varies more for better than for worse writers. The following "frequencies of significant variations" were observed in the groupings of students:

	Top 20	Bottom 20
Using the highest score:	36	8
Using the average score:	22	11
	Upper 40	**Lower 40**
Using the highest score:	63	23
Using the average score:	55	31

Noting that use of the average score for ranking tended to reduce the frequency of significant variations, the investigator used the highest score of each student as the basis for his subsequent analysis of differences between better and worse writers.

Further observations based on the above data included these:

1. There were more than four times as many significant variations in the quality of writing of the top 20 than in the bottom 20.
2. Of the 50 variations of two or more letter grades, 43 occurred among the upper 40 students while only seven occurred among the lower 40.
3. Of the 31 students having variations of two or more letter grades, 25 were among the upper 40 students, only six among the lower 40. Of the 25 from the upper group, 15 had two or more such variations; of the six from the lower 40, only two had two or more such variations out of the four such variations possible.

As a result of such observations, the investigator concluded "that a *single paper* written by a student on a *given topic* at a *particular time* cannot be considered as a valid basis for evaluating his achievement in a writing course at any time, unless that student's writing ability was rather low; and, even then, a single paper would not provide an infallible basis for such an evaluation."

After examining the validity of evaluating student achievement in writing by a single paper, the investigator compared content factors (from different topics) and day-to-day efficiency as sources of variations in quality of writing. He tabulated the number of significant variations in quality of writing "*between days* on the same topics and *between topics* on the same days":

	Between Days	Between Topics	Combined Totals
Top 20	17	19	36
Bottom 20	8	7	15
Upper 40	35	28	63
Lower 40	12	11	23
Total Group	47	39	86

If content factors and day-to-day efficiency were equally responsible for significant variations in quality of writing for the total group, the 86 variations would be divided into 43 for each factor. But application of the chi-square test indicated, at the 5 percent level of confidence, that the observed frequencies of 47 and 39 were not significantly different from 43. On the other hand, the differences observed between the top and bottom, and between upper and lower students, were significant at the 1 percent level. Hence the investigator concluded that content factors and day-to-day efficiency "were responsible for comparable variations in the quality of student writing" for the total group but not for comparable variations in the quality of writing by high and low students.

The above tabulation for "between topics" included both similar and dissimilar topics. To ascertain whether or not these types of topics have comparable effects on the quality of student writing, the investigator made a new tabulation of the significant variations observed "between topics":

	Dissimilar Topics	Similar Topics	Combined Totals
Top 20	9	10	19
Bottom 20	6	1	7
Upper 40	15	13	28
Lower 40	10	1	11
Total Group	25	14	39

If similarity and dissimilarity of topics were equally responsible for significant variations in quality of writing for the total group, the 39 variations would be divided into 19½ for each set of topics. But application of the chi-square test indicated, at the 5 percent level of confidence, that the observed frequencies of 25 and 14 were not significantly different from 19½. However, although the differences observed between similar and dissimilar topics was negligible for the top 20 and upper 40 groups, it was significant for the lower 40 students. Hence, the investigator concluded that although both the similarity and dissimilarity of topics were about equally responsible for variations in the quality of writing by better writers, only the dissimilarity of topics was responsible for significant variations in the quality of writing by worse writers. (The reader should be reminded here that the "similar topics"– Topics 1 and 2–both demanded a sort of low-level argumentative mode of discourse, and Topic 3–involved in the pairing of "dissimilar topics"–demanded the expository mode. In other words, the variations in the quality of writing of worse writers may result more from a change in mode of discourse than from a change in subject matter. The subject matter of all three topics was strongly similar in that in each case it involved matters commonplace for Michigan State freshmen.)

Finally, in order to determine whether or not examination pressure had an adverse effect on student writing, the investigator tabulated the gains and losses in the scores of the "pressure" and "nonpressure" students from the first day of writing to the second.

Nonpressure Groups A and B	Gains	Losses	Combined Gains and Losses
Upper 40	6	11	17
Lower 40	3	2	5
Total Group	9	13	22
Pressure Groups C and D			
Upper 40	10	7	17
Lower 40	5	3	8
Total Group	15	10	25
Grand Total	24	23	47

Although more losses than gains were observed among the nonpressure groups and more gains than losses among the pressure groups, application of the chi-square test indicated, at the 5 percent level, that the differences could be attributed to chance variations. The frequencies of significant gains and losses for the lower half of the students were too small to permit statistical analysis, but no pronounced trends were observed. In short, the investigator concluded that "the psychological pressure of the examination is not likely to have an adverse effect on the quality of writing by a number of students."

In summary, where a *group* of college freshmen is concerned, the *average* quality does not seem to be affected by variations in the day-to-day efficiency of individuals or by the pressure of the examination situation. Although the study indicated that the average quality was not affected by the nature of the topics assigned, there is basis for discounting this conclusion in favor of the notion that variation from expository to argumentative mode of discourse does not seem to affect the average quality of the writing of a group of freshmen.

On the other hand, variations in the quality of writing of *individual* freshmen did seem to result from variations in the day-to-day efficiency of students, especially of the better writers, but the pressure of final examination conditions did not appear to affect adversely the writing of either the better or the worse writers. Although the writing of the better writers did not seem to vary significantly in quality according to whether the students were writing expository or argumentative papers, the quality of writing by the worse writers did reveal a significant number of variations according to mode of discourse.

Among the implications which the investigator drew from his findings were these:

1. The justification seems very doubtful for grading a college freshman's achievement at the end of a writing course on the basis of one writing sample, especially for better writers.
2. An evaluation of the overall, or average, group improvement resulting from a writing course may be obtained from a single pretest theme and a single post-test theme.
3. An evaluation of "the effectiveness of different teaching methods with students who vary in writing ability" should be based on several samples of writing from each student, writing on different topics (modes of discourse) and during different days, the same procedure being followed both with the pretest and the post-test themes, the nature of the assignment being similar if not the same on both these occasions.
4. The post-test papers may be used as a final examination without seriously affecting the results.

The Smith Study

Smith, Dora V. *Class Size in High School English: Methods and Results.* Minneapolis: University of Minnesota Press, 1931. 309p.

Problem:

The purposes of this study, conducted in the ninth grade at the University High School of the University of Minnesota, were "(1) to discover the effect of the size of the class upon the efficiency of instruction in ninth grade English and (2) to devise techniques whereby a large group may be handled effectively with a minimum of waste of time and activity and a maximum of attention to and response from the individual." Although the investigation was conducted during 1925-1927, the problem and procedures are meaningful today, especially in view of current explorations in the use of lay readers to help English teachers handle composition marking for large classes. In this study, however, the investigator, who taught the large and small classes involved, had "no assistance in the actual conduct of the class nor in the handling of individual pupils," including the marking of compositions. Moreover, although it is not explained in the published report, the investigator was not concentrating most of her efforts on teaching the experimental and control classes: in addition to them, she was teaching a class of senior English with 32 pupils, sharing substantially the teaching of a two-hour graduate seminar containing 21 experienced teachers, supervising the work of nine student teachers, acting as adviser to the freshman class, and chairing a faculty curriculum committee. In short, the study is a report of what can be achieved with a large class by a teacher generally known to be highly proficient and energetic but teaching with at least a normal instructional load.

When the study was being considered for inclusion in this report, a question was raised about how representative the University High School ninth graders would be of school pupils in general. Her published study does not bring out this point, but the investigator has informed the writers of this report that the ninth graders had come from thirty different Minneapolis schools for this, their first year in the University school. The fact that pupils applied for admission to the University High School, where undoubtedly tuition was charged, and the pupils' average IQ of 116 suggest that the investigation was conducted with groups of superior intelligence and socioeconomic background. Consequently, the conclusions of this investigation have relevance to superior ninth grade pupils rather than to ninth graders representative of the American school population as a whole.

This experiment was carried on and evaluated for two years, but many of the techniques for dealing with the large class were tried during the first year for the first time. Furthermore, the teaching included literature and the mechanics of reading as well as written composition. The following summary will, of course, focus on the composition aspects of the experiment during the second year. In some matters, however, this focus cannot be maintained; sometimes statistics were reported only for the two-year period in such a manner that the data for the second year cannot be separated from the data of the first year.

In addition to the above qualifications, the reader's attention is directed to the complex but careful procedures used in matching the pupils in this experiment. Concerned throughout with English in general, not solely with written composition, this study nevertheless investigated the composition in a total English class setting with a care which merits attention today even though, over thirty years later, the reader will desire additional studies of class size which may corroborate or refute the findings of this investigation.

Procedures

Two ninth grade classes were used in the 1926-1927 school year, a control group of 21 pupils and an experimental group of 51. In the large class, 21 of the pupils were matched with the pupils of the small class; the other 30 pupils in the large class were included to build up the size of the class.

Matching was done on the basis of a battery of tests. A number of intelligence tests were administered at the beginning of the school year: the Army Alpha, the Haggerty Intelligence Examination, Delta II, the Pressey Senior Classification Test, and the Terman Group Test of Mental Ability, Form A. The median IQ for each pupil was then used in pairing. The pupils were also paired for chronological age, reading ability (using the Haggerty Reading Examination, Sigma III, Form B), grammar (using the University High School Grammar Test, "virtually a pretest in the grammar portion of the course in ninth grade English"), and spelling (using "a pretest in the ninth grade minimum essentials spelling list for the state of Minnesota"). In addition, the matching was

Table 1

Pretest Results for Paired Pupils, 1926-1927

	Range	Mean	Standard Deviation
Average I. Q.			
Small	96-136	115.90	10.53
Large	97-142	116.43	10.62
Chronological Age			
Small	157-186	167.43	7.26
Large	155-186	167.14	9.20
Reading			
Small	65-128	96.67	17.09
Large	75-128	96.71	14.28
Grammar			
Small	30-90	63.38	17.15
Large	30-85	61.86	13.93
Composition			
Small	3.4-6.7	5.19	.78
Large	4.4-6.5	5.36	.63
Spelling			
Small	69-133	118.81	14.06
Large	97-134	118.00	8.83

based on an initial composition written and rated according to the instructions for the Hudelson Typical Composition Ability Scale, which requires the reproduction of a narrative ("Slatter's Hill Snowball Fight"). The compositions were rated by five teachers, none of whom knew which papers were written by which class (although that was not made clear in the published report). The average score was used in matching the pupils. No data were offered on reader reliability, but the investigator reported that "all tests used in the experiment were rescored by an outside scorer before any computations were made." Scores for each pupil on each pretest are given in the published report, from which the figures in Table 1 are taken. The investigator further reported that "In no case did the difference exceed the probable error of the means except in composition. From the standpoint of significance, the chances that an equal difference would occur outside of these limits were even for chronological age, reading ability, grammar, and spelling. The chances for the differences in IQ and composition were 1.18 to 1." Thus, in matching pupils for all of the aspects of English studied, the large group was slightly favored with respect to composition.

The course of study, "identical in both classes," consisted of the ninth grade units usual for the University High School at that time:

> *First Quarter.*—Minimum essentials of English grammar and spelling; a short story unit with Rosa Mikel's *Short Stories for English Courses* as a starting point; composition and the monthly publication of a freshman paper.
>
> *Second Quarter.*—A unit on punctuation; minimum essentials of grammar (continued); the *Merchant of Venice* and other plays; composition and the continuation of the freshman paper.
>
> *Third Quarter.*—A unit in letter writing; a project in the literature of chivalry with *Ivanhoe* as a basis; a poetry unit based upon Spaulding and Spaulding's *Open Gates;* composition and the continuation of the freshman paper.

Although the investigator taught the two classes herself, she had three observers sit in each class daily and record "pupil participation in each class, the attitude of teacher and pupils, the physical features of the room, the time consumed in movements to and from the blackboard or in small-group activities, and the time devoted to each section of the lesson in each class." (The instructions given to observers are reproduced on pages 278-279 of the published report.) "The observers, all of whom were university seniors training to be English teachers, were shifted every three weeks, at the end of which they summarized their observations." The summaries, taken as a whole, thus "approximated a stenographic report of each day's lesson" throughout the year.

Rather than attempt to employ the usual methods of instruction in both classes, and thus handicap the large class with ill-adapted procedures, the investigator devised a number of techniques for use with the larger group, reporting them in Chapter VII. Sometimes the concomitant procedures for the small group were described or implied in the published report; sometimes they were not.

The large class was divided into six groups, each of which was drawn from the complete range of ability available and was seated together in the classroom. The locations were shifted each six weeks to equalize opportunity for favorable seating. Each group had an especially capable pupil as leader responsible for taking the attendance and collecting and distributing the papers of the eight or nine pupils in his group. Various other techniques were employed to facilitate the teaching of literature, reading, and speech to the large class.

When the large class had written compositions, the investigator had the pupils meet in their groups to read their papers aloud, select the best one to read aloud to the class, and make suggestions to each other for improving their writing. She noted that "every pupil was able to read his composition aloud . . . and, frequently on the same day, the pupils chosen could also give theirs before the class as a whole." A list of questions, mostly concerning rhetorical considerations but also sentence structure and mechanics, was given to the pupils before they wrote to guide them in their selections and suggestions. (See pages 157-158.) The investigator attested that this type of prevision was done with especial care before the pupils wrote business and social letters, and "So far as possible, letters were at least begun in class under the direction of the teacher."

The pupils also drilled or recited on grammar problems in their groups, limiting the material "to those portions of English grammar that, in the absence of scientific evidence, are believed to be functional in correct speech and writing." The investigator carefully instructed the pupils, when noting that a fellow group member had erred, "to question him in regard to the tests he applied before making his response," helping him to "think out" the correct response.

With the large class, the investigator found it more necessary than ever to break the 50-minute class period into three or more different activities and not to spend the entire time on "a procedure of one sentence each around the class." She reported that she had to go into the classroom with "a definite program that permitted a maximum of participation for the pupils," a definiteness which was "especially necessary if she was to make quick transitions from one part of the lesson to another with a minimum of confusion and waste of time." Still it was not possible, of course, to have as much pupil participation in the whole large class as in the small class. While fourteen lessons in grammar were being held during October, an actual count of pupil recitations was kept ("answering or asking an oral question during regular class discussion or volunteering some comment thereupon"). The recitations averaged 1.55 per pupil per day in the large class, 2.85 in the small class, with only one pupil in the large class reciting as often as the average pupil in the small class. The investigator reported that if the questions were to be distributed fairly in the large class, it was necessary to call on pupils most of the time and "pay little attention to volunteers."

Rather than have the large class recite as one group, the investigator often used the group recitation technique described in the paragraph before the preceding one. Then each pupil recited three or four times in 15 or 20 minutes, covering the two dozen or so sentences assigned for homework. The investigator went from group to group, settling disagreements and noting which three or four sentences caused enough problems to discuss with the pupils when the class reconvened as a whole. After this entire procedure (which took from 20 to 25 minutes), the investigator often administered a brief test before turning to new work. She remarked on the increased attention, interest, training in group cooperation, and intergroup competition but made no comment on whether or not the good pupils suffered from continually helping the poor pupils in this procedure. Somewhat similar group procedures were used in seeking more descriptive adjectives to use in writing.

The large class responded with more zeal than the small class to "anything in the nature of a contest," especially a spelling contest. To prevent the possibility of "harm to a nervous child with a special disability in spelling," the investigator saw to it that "No lesson . . . contained more than ten words selected from the minimum essentials list for the ninth grade in Minnesota."

During a two-week letter writing unit, the groups became business firms, each with a manager, overseer, and messenger having his own special duties. Although the pupils occasionally marked and corrected each other's papers, the investigator always gave the final score. The group envelope contained the list of pupils, the types of letters written, and a system of accounts: 10¢ for an A letter, 8¢ for a B, 6¢ for a C, 4¢ for a D, a 4¢ deduction for each late letter, and a dollar bonus for each firm having every letter finished on time. Each pupil wrote one or two letters each day during this unit. The amount of paper correction was so heavy at this time that the investigator could not return to each pupil in the larger group all his letters the day after he wrote them.

Various other composition projects were pursued in both classes, but they did not seem different because of the size of the class except in one respect: when compositions were being chosen to be read at a holiday program, more good stories could be found in the larger class but that number was still a smaller proportion of the large class than the proportion of papers used from the small class.

A major problem was providing for individual differences in the large class. One technique was to have each pupil record his scores from quarter to quarter on examinations on various aspects of the work, stimulating self-competition. Another was to group pupils for drill according to their needs, the investigator working with weaker pupils drawn from each group while the others proceeded as usual under the direction of their regular leaders. A third device was to seat in the front row the pupils who had experienced difficulty the previous day and to call on these pupils for review questions. Volunteers were occasionally permitted to take these seats to benefit from the review—and to help the class to welcome the device. The very nature of the composition work permitted individual tastes to operate in the selection of topics and individual needs to be met in the marking of papers. At the end of each quarter, the investigator had individual conferences with the pupils, reviewing their compositions (and an error check sheet occasionally used) and pointing out constant repetitions of errors. Pupils with similar repetitions were assigned to the same row of seats for laboratory work, permitting the investigator to give them special help.

Brief tests and exercises were given very frequently to compensate for less opportunity to participate in the large class. When occasionally such tests were written on the blackboard, the investigator filled the blackboard with writing large enough for all pupils to see; however, the constant looking up and down at the blackboard seemed to introduce errors not found when tests were mimeographed and was conducive to wandering glances at other pupils' papers. The pupils corrected most of these tests themselves and took interest in tabulating class errors and suggesting the next steps in teaching. The pupils also corrected each other's homework exercises in punctuation.

Results

The results of the experiment were tested at the end of the year by a number of measures, but only those results relating to composition are given here. Only the matched pupils in the large class are considered in comparison to the pupils in the small class.

Five objective tests were given in capitalization, punctuation, and spelling. The large class was slightly superior to the small one in capitalization and more clearly so in spelling, but the two classes were practically even in punctuation. Another five objective tests were administered to measure results in grammar, with no clear difference in favor of either class.

The investigator reported another measure which, at first examination, seems to be a more valid test of composition ability than the objective tests—mean total points for the approximately 30 papers per pupil written during the letter writing unit. Unfortunately, however, the investigator herself was the sole rater of the letters and the ratings of all the letters were used (not merely several of the latest for each pupil, reflecting the end result of differentiated instruction), largely discounting the significance of her finding that the small class was distinctly superior in this measure.

Doubtless the most valid measures of composition ability were the Hudelson Typical Composition Ability Scale (reproduction of a narrative) and the Hudelson English Composition Scale (original narrative on "The Most Exciting Ride I Ever Had"). The scores on each composition were "the average of five scale ratings for each paper given by five experienced judges." As with the narrative pretest scores (although it was not stated in the published report), the raters did not know which class the writer of a paper was in. No figures on reader reliability were afforded. It was stated that "two quarters intervened between the giving of test compositions" (evidently pretest and post-test compositions), affording a limited comparison of the composition abilities of the two groups.

In order to check the results of the study from another view, the investigator also examined the records of individual pairs matched at the beginning of the experiment. Here the investigator based her analysis on the combined results for the two years. On the objective testing in capitalization, 41 percent of the pupils in the large classes surpassed their mates, while but 34 percent of the pupils in the small classes surpassed theirs, 68 percent of the pupils in the large classes reaching or exceeding the mean of the small classes. These and other results are shown in Table 2. A slightly higher proportion of the pupils in the small classes surpassed their mates, but the pupils in the large class who surpassed their mates did so by a wider margin, especially in the reproduction of a story. It seems clear that, considered in their respective groups, these superior ninth grade pupils showed little difference in their proficiency at writing narrative compositions as a result of being in a larger or smaller English class when concerted efforts were made by a superior teacher to assure individual participation and attention in the larger class.

In addition to measuring the writing ability of the pupils, the investigator explored their attitudes toward small class and large class instruction. Six months after the experiment had begun, and without advance warning which would have permitted them to be influenced by each other or their parents, they were given a questionnaire to learn their initial and present reactions. The pupils were told that the questionnaire was being given to provide data for another investigator, and one of the pupils collected the completed papers and took them to his office without examination by the teacher.

Although the investigator noted that "the experiences of the 1925-1926 group may have influenced the attitude of incoming students the next year," such an influence doubtless would exist in any school where large classes were used year after year. Consequently, the results only for the 1926-1927 group are given here (when they can be isolated), and, since the published study reports for all pupils (not merely the matched pupils) in the large group, the reactions of the small class are compared to those of the entire large class. Because of absences on the day the questionnaire was filled out, only 46 pupils in the large class and 17 in the small class are included.

When asked on the questionnaire which class they had hoped to be placed in the previous September, 24 percent of the pupils in the large class indicated a remembered preference for the large class, 50 percent said they would have chosen the small class then, and 26 percent voiced no preference. After six months of experience in the class, 43 percent of the large class preferred it, 34 percent would still choose the small class, and 22 percent indicated no preference. The pupils whose preference had shifted from the small to the large class gave as their chief reasons "(1) that they learned just as much, (2) that they heard more opinions expressed and shared more varied talents, (3) that, contrary to their expectations, they received just as much help in the large class in English as they did in the small classes in

Table 2

Relative Progress of the Mated Groups in Composition and the Mechanics of English 1925-26 and 1926-27 Combined

Test	Average Percentage Possible Gain		Number Surpassing Mate		Percentage Surpassing Mate		Average Numbers of Points by Which Mate Is Surpassed		Number at or Above Mean of Other Group		Percentage at or Above Mean of Other Group	
	Sm*	Lm*	Sm	Lm	Sm	Lm	Sm	Lm	Sm	Lm	Sm	Lm
1. Pressey Capitalization	.....	7.36	14	17	34	41	28.97	41.33	19	28	46	68
2. Pressey Punctuation	14.22	.....	27	13	66	32	28.54	23.07	29	15	71	37
3. Briggs English Form	.....	19.14	6	15	29	71	56.66	51.60	8	16	38	76
4. University High School Punctuation (scores)†	.54	.....	26	14	63	34	7.37	11.44	26	20	63	49
5. Spelling (1926)	10.52	.....	10	11	48	52	55.10	34.00	17	15	81	71
6. Kirby Grammar-Sentences†												
1925 scores	.40	.....	10	8	50	40	3.80	3.75	13	9	65	45
1926 percentage of possible gain	2.53	.....	24	15	59	37	35.06	50.38	30	20	73	49
7. Kirby Grammar-Principles†												
1925 scores	1.40	.....	9	8	45	40	6.33	3.63	13	8	65	40
1926 percentage of possible gain	2.62	.....	19	19	46	46	48.10	28.73	23	19	56	46
8. University High School Grammar	.....	.22	23	17	6	41	17.71	25.09	22	22	54	54
9. Minnesota English Essentials	.....	.63	20	21	49	51	20.03	21.56	23	26	56	63
10. Wilson Language Errors	.17	.....	18	17	44	46	3.25	2.90	21	23	51	56
11. Letter Writing (scores)†	25.61	.....	24	17	59	41	53.40	32.90	25	11	61	27
12. Hudelson Typical Composition Ability	2.98	.....	22	17	54	41	18.41	31.51	23	20	56	49
13. Hudelson Exciting Ride	.....	3.63	22	18	54	44	26.64	27.10	23	18	56	44

*Sm = small class mated; Lm = large class mated.

†Where scores were used one year and percentage of gain the next, average differences are given separately, but numbers surpassing mates and so forth are combined. In the University High School Punctuation Test and the unit of letter writing, scores only were available.

other subjects, and (4) that they became acquainted with more members of the freshman class." (The third reason, of course, is suspect here; we have no evaluations of the teaching proficiency of the teachers in other subjects—whose abilities may not have equalled the investigator's—and hence we cannot attribute the help necessarily to the large class per se.) Of the reasons given by the large-class pupils for disliking the larger group, the chief were that there was (1) less "opportunity for participation and individual attention and (2) more noise and confusion from many pupils working in the same room at the same time." Four students missed a feeling of intimacy within the larger group. Only one student referred to difficulty hearing during the group activities, although more mentioned such a problem when the class met as a whole. Thirty-five percent of the large class preferred that class specifically for composition work, but 9 percent preferred the small class for that activity; over half of the large class evidently did not think the difference important enough to mention.

The investigator also studied the responses of her observers to twelve questions each was asked to answer. When asked in which class the teacher and pupils knew each other better, 65 percent of the observers stated that there was no difference and 20 percent noted that, although there was no difference between teacher and pupils, the pupils in the large class did not know each other as well as those in the small class knew each other. Although one-fourth of the observers noted more shyness in the small class, one-third saw none in either class, and one-third saw the same amount in both. Forty percent of the observers thought the pupils in the large class asked more questions, but 60 percent indicated that the number of questions was equal in the two classes. The most striking agreement among the observers related to the spirit of each class, four-fifths making a strong endorsement of the superior interest and enthusiasm of the large class, one-fifth noting that the spirit of the two classes was the same. Citing the difficulty of summarizing evidence on the discipline in the two classes "because of varying views as to what constitutes 'good discipline,'" the investigator nevertheless offered her own observation that "Restlessness and inattention *when the class met as a unit for general discussion* were on the whole more prevalent in the large class than in the small one, and, once started, were harder to quell." Doubtless it was not entirely such situations, however, which led all of the observers but one to state emphatically that the large class was more difficult than the small one to manage, even though the investigator herself affirmed that she "entered the large class with more of a feeling of responsibility and with a greater sense of adventure than she did the small one." The investigator's candid revelation of such feelings may lead some readers to believe that her objectivity in reporting was extended to an objectivity in treating both classes the same insofar as possible; other readers, however, may believe that a natural interest in the success of the large class would contribute, even with a person as fair-minded as this investigator, to the Hawthorne effect —the tendency of any experimental group to surpass its usual efforts because an experiment heightens its interest in learning.

Finally, the investigator made some of her own observations on the teacher's load and the expense to the school system. She thought that it did not take her quite two and a half times as long to correct 50 papers as to do 20, or to plan "group activities and assignments where a topic is arranged for each individual member of the group." She also noted that a teacher would seldom wish to meet a class of 50 without special planning, though he might occasionally do that with but 20 pupils in a class. Although she paid tribute to the value of "great quantities" of mimeographed materials in the large class, she also stated that, when not in an experimental situation necessitating much testing, she probably could dispense with half of what she used and, at much less expense than the salary of another teacher, obtain clerical help in checking diagnostic check sheets and objective tests. In fact she estimated, using the pay rates of 1926, that a school could save 28 percent of the teacher expense if it assigned teachers three classes of 50 pupils each and clerical assistance for three hours a day rather than five classes of 20 pupils each and no clerical assistance, but she questioned "how many classes of 50 a teacher may teach in one day without harm to herself or to the pupils."

In conclusion, the Smith study, although probably affected by the over-involvement of this obviously remarkable teacher-investigator with the actual conduct of the experiment, still offers strong indication that the teaching of composition to superior ninth graders may be generally as effective in large classes as in small classes if the teacher has the requisite ability and is given adequate time for preparing lessons and adequate assistance in handling the routine aspects of class management. Whether large class teaching may be further made practicable for more typical teachers and pupils by the judicious use of competent lay readers and modern devices like the overhead projector remains to be demonstrated in a carefully controlled experiment. But when further investigations are made of the effectiveness of teaching composition to large classes, the teaching techniques and experimental procedures of the Smith study will offer many suggestions and a few noteworthy cautions.

The Becker Study

Becker, Samuel L., *et al. Communication Skills: An Experiment in Instructional Methods.* Iowa City: University of Iowa, 1958. 43p. The essential information of this publication and some additional analyses may be found in Samuel L. Becker and Carl A. Dallinger, "The Effect of Instructional Methods upon Achievements and Attitudes in Communication Skills," *Speech Monographs,* XXVII (March, 1960), 70-76.

Problem

In this University of Iowa freshman program* integrating the teaching of reading, writing, speaking, and listening with a staff largely composed of part-time graduate assistants, an experimental comparison was made in 1957-58 of the effectiveness of three methods of teaching: (1) the Normal method, making one instructor responsible for the training of his students and meeting them in class four times each week, (2) the Bibliography method, making the student more "self-reliant and independent by reducing from four to three the number of formal class meetings per week, by presenting the basic principles of the course . . . in a bibliography of assigned and collateral readings, and by devoting most of the class hours to performances under the guidance and criticism of a classroom instructor," and (3) the Kinescope method, presenting "the basic principles of the course by 'experts' through the use of film recordings of television programs (kinescopes), supplemented by discussion of these principles and their application in performances under the guidance and criticism of the regular classroom instructor." The Kinescope classes met four times each week, at one meeting of which a 29-minute kinescope usually was shown and discussed. Essentially the same material was presented under each method. (See the Schedule of Kinescopes on a later page.)

As he reads this summary, the reader is invited to note the careful statistical procedures used and the possibility that results which are not statistically significant may be meaningful for that reason. He is also advised to observe that some of the unsatisfactory aspects of this experiment might have been prevented if a pilot investigation had preceded the regular study.

---

*The name of the program was changed from the Communication Skills Program to the Rhetoric Program in 1961.

Procedures

All three methods were used in the regular two-semester course (10:1 and 10:2), but only the Normal and Bibliography methods were used in the accelerated one-semester course (10:3). Classes were apportioned equally among the methods at each hour, and students were randomly assigned to the classes at the hour for which they registered. Except that "instructors teaching more than one section of the course were assigned to classes all within one treatment group" and "an attempt was made to achieve a balance of experienced and inexperienced instructors," instructors were randomly assigned to classes. If a class did not have the some instructor in both 10:1 and 10:2, it was eliminated from the experiment. If an instructor taught more than one section, only one was counted in the experiment, facilitating control of instructor differences. If complete pretest and post-test scores could not be obtained for a student, his scores were eliminated from consideration. In the final data, there were 758 students represented, 585 in the two-semester course and 176 in the one-semester course, divided among the methods approximately as explained above.

Detailed, day-by-day syllabi were distributed to the instructors and students of the Bibliography classes (10:1, 10:2, and 10:3) and the Kinescope classes (10:1 and 10:2). These five syllabi were prepared by the supervisors of the course, guided by the suggestions of Communication Skills staff committees which met during the spring and summer preceding the year of the experiment. (Although it is not reported in the published study, these syllabi were not distributed to the staff until a few days before instruction began, making it difficult for the instructors to familiarize themselves with the course plan or the bibliographic references.) The Bibliography syllabi included the bibliographies of collateral readings on the topics covered in the course. At least one copy of each reference was placed on "reserve" circulation at the University Library. "Methods of stimulating these sections were left to the instructors of these classes." Excerpts from the syllabi are given in Appendix B in the full report. In 10:1, the emphasis was on expository writing and speaking, and reading and listening to exposition. In 10:2, the emphasis shifted to argument and, to a lesser extent, criticism. In 10:3, the emphases of both 10:1 and 10:2 were covered. The textbooks were the same for all three courses:

Baird, A. Craig, and Franklin H. Knower. *Essentials of General Speech*. New York: McGraw-Hill Book Company, 1952.

Blair, Walter, and John C. Gerber, *Factual Prose*, 3rd ed. Chicago: Scott, Foresman and Company, 1955.

Garner, W. L., D. G. Pugh, and D. E. Wylder. *Reading Factual Prose*. Chicago: Scott, Foresman and Company, 1955.

McCrimmon, James M. *Writing with a Purpose*, 2nd ed. Boston: Houghton Mifflin Company, 1957.

Twenty-eight half-hour kinescopes were used in the kinescope classes, all but two prepared for this purpose at the University of Iowa. Simultaneous showing of the kinescopes in two different classrooms was made possible by printing two copies of each film. No more than three classes (with a maximum of 25 students per class and an average less than 24) gathered in each screening room for showings, and after the film was shown they returned to their regular classrooms for discussion of the topic. (Although it is not mentioned in the published report, it is true that the instructors rarely previewed the kinescopes before their students viewed them. Before some kinescopes were shown, the instructors were supplied with outlines of the content but rarely with suggested questions to use in guiding class discussion. It is not surprising, then, to learn that a few instructors considered the kinescopes little more than an interruption of their regular teaching.) The topics of the kinescopes ranged from the theoretical to the practical, but the larger proportion was practical. Some 16 of the kinescopes may be said to concern communication in general or to relate to speech as well as to writing; 5 relate directly to writing; 7 relate mostly to speaking, reading, or listening. The schedule for the kinescopes follows:

Schedule of Kinescopes
1957-58

Topic	"Expert"	Date Shown
"The Nature of Communication"	Dr. Harold C. Shiffler, Assistant Professor, Dramatic Art, SUI	Oct. 2, 1957
"Ideas in Communication"	Dr. Orville A. Hitchcock, Professor, Speech, SUI	Oct. 8, 1957
"Framework of Ideas"	Dr. Carl A. Dallinger, Associate Professor, Speech; Coordinator of Communication Skills, SUI	Oct. 15, 1957
"Adjustment to Communication"	Dr. Robert L. Gregg, Assistant Professor, Speech, SUI	Oct. 22, 1957

Topic	"Expert"	Date Shown
"Concepts of Purpose"	Dr. John C. Gerber, Professor, English, SUI	Oct. 29, 1957
"Bodily Action"	Dr. Carl A. Dallinger, Associate Professor, Speech; Coordinator of Communication Skills, SUI	Nov. 1, 1957
"Paragraphing"	Dr. Clark Griffith, Assistant Professor, English, SUI	Nov. 12, 1957
"Basic Patterns of Organization"	Dr. Harry H. Crosby, Assistant Professor and Writing Supervisor, Communication Skills, SUI	Nov. 19, 1957
"Effective Reading"	Dr. William Eller, Associate Professor and Director of Reading Laboratory, Education; Reading Supervisor, Communication Skills, SUI	Nov. 26, 1957
"The Support of Ideas"	Dr. John C. Gerber, Professor, English, SUI	Dec. 3, 1957
"Research Techniques"	Miss Cleo Martin, Instructor, Communication Skills, SUI	Dec. 10, 1957
"Effective Sentences"	Dr. John C. McGalliard, Professor, English, SUI	Dec. 13, 1957
"Documentation and Manuscript Techniques"	Mr. William J. Holmes, Instructor, Communication Skills, SUI	Jan. 7, 1958
"The Process of Revision"	Dr. Clark Griffith, Assistant Professor, English, SUI	Jan. 10, 1958
Second Semester, 10:2 Classes		
"Difference Between Words and Things"	Dr. Irving J. Lee, Professor, Speech, Northwestern University	Feb. 13, 1958
"Voice in Communication"	Dr. Harold C. Shiffler, Assistant Professor, Dramatic Art, SUI	Feb. 18, 1958
"Why People Misunderstand Each Other"	Dr. Irving J. Lee, Professor, Speech, Northwestern University	Feb. 25, 1958
"Articulation and Pronunciation"	Dr. Harold C. Shiffler, Assistant Professor, Dramatic Art, SUI	Feb. 28, 1958
"Making Language Appropriate and Effective"	Dr. Clark Griffith, Assistant Professor, English, SUI	March 7, 1958
"Mass Media as Background in Communication"	Dr. Richard Braddock, Assistant Professor, Communication Skills, SUI	March 18, 1958

Topic	"Expert"	Date Shown
"Obstructions to Critical Listening and Reading"	Dr. William Eller, Associate Professor and Director of Reading Laboratory, Education; Reading Supervisor, Communication Skills, SUI	March 20, 1958
"The Argumentative Situation"	Dr. Orville A. Hitchcock, Professor, Speech, SUI	March 25, 1958
"Evidence, the Foundation of Reasoning"	Dr. Orville A. Hitchcock, Professor, Speech, SUI	April 1, 1958
"Processes of Reasoning"	Dr. Carl A. Dallinger, Associate Professor, Speech; Coordinator of Communication Skills, SUI	April 10, 1958
"Discussion, Its Basis and Forms"	Dr. Robert F. Ray, Professor and Director, Institute of Public Affairs, SUI	April 15, 1958
"Advocacy, Its Basis and Forms"	Dr. Orville A. Hitchcock, Professor, Speech, SUI	April 22, 1958
"Common Fallacies"	Dr. William Eller, Associate Professor and Director of Reading Laboratory, Education; Reading Supervisor, Communication Skills, SUI	April 29, 1958
"Communication Which Evaluates and Criticizes"	Dr. John C. Gerber, Professor, English, SUI	May 13, 1958

Although the report of the experiment was concerned with reading, writing, speaking, and listening, only the writing will be treated in detail here.

The pretests and post-tests each included a 45-minute objective test in English composition and a 450-word minimum, two-hour expository theme. The themes were rated from 1 to 7 on each of these categories: purpose, content, organization, sentences, diction, and mechanics. (An analysis of each of these categories is given on the Theme Book cover, reproduced on page 40 of the original report, and samples of the instruction sheets given to the students are reproduced on pages 38-39.) The six ratings were then added to yield a total score. Pretest themes were rated by one experienced member of the Communication Skills staff. The 10:3 post-test themes were rated by three members; the 10:2 post-test themes were rated by two members with another reader added in cases of marked discrepancy. The final rating was the mean of these ratings. In each theme test, the student was given a choice of four topics —different on each testing occasion: "The Cold War," "Profit from Friendship," "My High School," etc., for the pretest; "The Significance of an Incident in History," "Special Knowledge I Possess," "Family Relationships," etc., for the post-test. (Although it is not reported in the published study, the principal investigator has informed the writers of this report that the reliability of the rating was .83.)

In addition, several attitude tests were administered to students and instructors. At the beginning and conclusion of the experiment, a "Study of Educational Attitudes" test was administered to the students apart from their Communication Skills classes and dissociated from the course. This test included 66 multiple-choice items asking how important the student felt various aspects of the communication skills were, and, interspersed among them, 12 similar items about other college subjects. The student was also asked to rate himself on each of the skills of communication and to reveal certain matters about himself, his family, his interests, and his experiences.

Directly associated with the Communication Skills classes was a questionnaire administered near the close of the course, inquiring which of the three methods the students preferred, which they believed easiest, and which most effective. A similar questionnaire was filled out by the instructors.

Results

The investigators reported that before beginning the analyses of data they decided to use the 5 percent level of confidence as the level at which to reject the hypothesis of no difference.

Simple analyses of variance were made to test the homogeneity of the 10:1 and 10:3 classes on each of the pretests and the placement test Composite Score (indicating the percent of students the sum of whose percentile ranks on the English, Vocabulary, Reading Comprehension, and Mathematics tests falls below the student's sum—a score which had been "found to correlate highly with academic achievement in college"). The classes did not differ significantly on any of the pretest measures considered here.

A groups-within-treatments analysis of covariance design, with the individual as the unit of measurement, was planned to test the differences among methods of instruction. (This results in a more precise analysis when individual scores are used.) But before making that analysis, a test for differences among classes within each method was conducted to determine whether or not there were extreme or systematic differences among instructors.

Analysis of covariance did show significant instructor differences for both the objective English composition post-test and the theme post-test in the two-semester course, suggesting the possibility that 10:1-10:2 instructors with large or small classes would carry disproportionate weight in the analysis. Consequently, the differences among methods of instruction were tested with a groups-within-treatments analysis of covariance with the section mean as the unit of analysis. The F-ratio on the objective English test was less than 1 and on the theme test was 2.64, where 3.30 is significant with 2/33 degrees of freedom; thus there was no significant difference in achievement indicated among the groups being taught by the three methods.

Analysis of covariance indicated significant differences among instructors only on the theme post-test in the one-semester course. The F-ratio for the adjusted groups-within-treatments mean squares divided by the adjusted within-groups mean squares was 1.49 for the objective English test and 2.07 for the theme test where F = 2.07 is significant with 2/166 degrees of freedom. When the theme data were tested with t-tests for differences between individual pairs of means, it was found that one of the Normal classes had a significantly higher mean than all other classes except another Normal class, which had a significantly higher mean than 5 of the other 8 classes. (There were 9 accelerated classes altogether—4 Normal and 5 Bibliography.) On the other hand, one of the Bibliography classes had a significantly lower mean than 6 of the other 8 classes, and another Bibliography class had a significantly lower mean than 4 other sections. To determine whether the difference between the Normal and Bibliography methods could be attributed to chance rather than true differences in the population, these theme data were grouped into two simple random analyses of covariance design, one with the individual score and the other with the section mean as the unit of analysis. The latter prevents instructors with larger classes from carrying more weight in the analysis. The F-ratio for the individual analysis for the adjusted treatments mean squares divided by the adjusted groups-within-treatments mean squares was 1.13 for the objective English test and 4.48 for the theme test, where F = 3.91 is significant with 1/173 degrees of freedom. This difference on the theme test favored the Normal method, but the investigators declined to accept this result as significant, noting that the difference may well have been the product of instructor differences when only four classes were included in the Normal accelerated group.

Analysis of variance utilizing a treatments-by-level design (four levels of approximately equal numbers of students, established by the Composite Score) was used to test whether or not students of different levels of academic aptitude achieved more or less in theme writing under each method. A significant interaction being found between levels and methods in both the two-semester and the one-semester courses, differences among methods of instruction were tested for each level independently for each course, and differences among levels were tested for each method independently for each course.

Within the next to the top level in the two-semester course, the Kinescope group averaged significantly lower theme scores than the Normal or Bibliography groups, which did not differ significantly. No significant differences among methods were found at any other level in the two-semester course. Although significant differences among some levels were found for each of the methods of instruction, the differences between pairs of levels were somewhat unsystematic in the two-semester course. In no case did a group at one level average a significantly higher theme score than a group at a higher level.

In the one-semester accelerated course, no significant interaction was found between levels and methods for the theme post-test, but significant differences were found among both levels and methods of instruction. The Normal group again scored significantly higher than the Bibliography group, and the top-level group scored significantly higher than each of the other levels. The other levels did not differ significantly among themselves.

On the "Study of Educational Attitudes" test, administered near the end of the experiment, analyses of variance showed no significant differences among instructors (classes) within each method or among methods of instruction.

When the students in the two-semester course filled out the questionnaire on their attitudes toward the three methods of instruction, on all three questions the Normal method was obviously the most favored among both the Normal students and the Bibliography students, while the Kinescope students clearly favored the method which they had experienced. The Bibliography method was least favored by the Normal and Kinescope students, but the Kinescope method was least favored by the Bibliography students. In the one-semester accelerated course, both the Normal and Bibliography groups favored the Normal method, the Normal students reacting overwhelmingly in that fashion. Summing up the results of the attitude questionnaire, the investigators reported, "Though students, in general, seem to prefer the Normal method of instruction, experience with either of the other methods tends to make them much more favorable to the method which they have experienced."

Near the end of each course, 44 instructors who had taught in the two-semester course and 8 instructors who had taught in the one-semester course filled out questionnaires indicating their attitudes toward the three methods. Although theoretical cell frequencies were not great enough to permit the calculation of a chi-square, it was very clear that these instructors preferred the Normal method of instruction and overwhelmingly rejected the Bibliography method. Their primary reason was that the Normal method provided the most opportunity for discussion, and many noted that the other methods afforded too little time to adjust to the individual needs of students. A few liked the standardization and better coverage of lecture material through the kinescopes.

In conclusion, there seemed to be no consistently significant difference in the effectiveness of the three methods in the two-semester regular course. The Normal method may have been more effective than the Bibliography method for teaching theme writing to freshmen in the one-semester accelerated course, but the observed difference may well have been the result of chance. It is possible that the Kinescope method would have fared better had lecturers more experienced with television made the kinescopes and that the Bibliography method would have fared better had the bibliographies been more selective. It is also possible that both the experimental methods would have been more successful if the instructors had had more experience with them, but the attitudes of the instructors seem to work against the likelihood. Although the Bibliography method (classes meeting one less time per week) permits the saving of classroom space and of instructor time in class, it may be less effective as a method of teaching composition to college freshmen who are generally of above average academic ability.

If the Becker experiment had been preceded by a pilot experiment —permitting the development of more satisfactory kinescopes and bibliographies and perhaps a more supportive attitude on the part of the instructors, the Kinescope and Bibliography methods might have been shown significantly superior to the Normal method. The real conclusion of the experiment is that—despite the factors working against both experimental methods—the Normal method was *not* significantly superior. That in itself is significant, in a different sense, because it suggests that college freshmen, perhaps especially those of average ability, can learn to improve their writing as well when new concepts are presented by film or by a bibliography of references as when new concepts are presented by individuals on a staff composed largely of part-time graduate assistants among whom there is a fairly high annual turnover.

V.

REFERENCES FOR FURTHER RESEARCH

Summaries and Bibliographies

A "classic" source to consult for research in the twenties and before is Rollo L. Lyman's *Summary of Investigations Relating to Grammar, Language, and Composition,* published in 1929 as a Supplementary Educational Monograph by the University of Chicago. Lyman gave summaries of more than 250 studies and offered many excellent suggestions for the improvement of research. Periodic summaries or bibliographies published in the intervening period include the following:

Encyclopedia of Educational Research. Published at ten-year intervals. See entries on "English—Language, Grammar and Composition," "Spelling," "Handwriting," etc. Usually draws heavily from next item.

Review of Educational Research. Summary of research on English composition included approximately every three years. Tends to include published research rather than dissertations and theses. Journal.

Bibliography of Research Studies in Education. Bulletin publication of the Office of Education, U.S. Department of Health, Education, and Welfare. Annual.

Elementary English. Yearly reviews of published and unpublished studies of the preceding year. Journal.

Elementary School Journal. Yearly reviews of references on instruction in elementary school English. October issue.

English Journal. Yearly reviews of research in secondary English, prepared by the NCTE Committee on Research.

School Review. Yearly reviews of references on instruction in secondary school English. February issue. Journal.

Since 1950, a number of bibliographies and summaries have been published in addition to the many helpful bibliographies included with other books, articles, and unpublished manuscripts. Some titles are offered here:

Brown, Roger. *Words and Things.* New York: Macmillan (Free Press of Glencoe), 1958. 398p.

Burrows, Alvina T., *et al. Children's Writing: Research in Composition and Related Skills.* Champaign, Illinois: National Council of Teachers of English, 1961. 73p.

__________. *What Research Says to the Teacher: Teaching Composition.* Washington: Department of Classroom Teachers and American Educational Research Association, National Education Association, 1959. 32p. Concerns elementary school children.

Carroll, John B. *The Study of Language.* Cambridge: Harvard University Press, 1953. 289p. Summarizes linguistic, sociological, and psychological studies of language.

Dawson, Mildred A. "Interrelationships between Speech and Other Language Arts Areas," *Elementary English,* XXXI (April, 1954), 223-233. Includes written communication.

Godwin, L. Ruth. "Studies Related to the Teaching of Written Composition," *Canadian Education and Research Digest,* III (March, 1963), 35-47.

Gray, William S. *The Teaching of Reading and Writing.* Chicago: Scott, Foresman and Company, 1956. 281p. An international survey of world literacy problems with reference to methods of teaching.

Haugh, Oscar M. "Representative Research in the Communication Skills," *Education,* LXXII (March, 1952), 470-480.

Horn, Ernest. *What Research Says to the Teacher: Teaching Spelling.* Washington: Department of Classroom Teachers and American Educational Research Association, National Education Association, 1954. 32p.

Hunnicut, C. W., and William J. Iverson, eds. *Research in the Three R's.* New York: Harper and Row, Publishers, 1958. 446p.

Hunt, Jacob T. "Selected Bibliography on Communication," *High School Journal,* XXXVII (January, 1954), 127-130.

Kraus, Silvy. *Teaching of Written Composition in the Public Schools: A Summary of Research.* Curriculum Bulletin, XV, 190. Eugene: University of Oregon, School of Education, 1959. 23p.

Leopold, Werner F. *Bibliography of Child Language.* Evanston, Illinois: Northwestern University Press, 1952. 115p. Includes titles through 1946.

Meckel, Henry C. "Research on Teaching Composition and Literature," pp. 966-1006 in *Handbook of Research on Teaching,* ed. N. L. Gage. A Project of the American Educational Research Association, National Education Association. Chicago: Rand McNally and Company, 1963.

Pavlak, Stephen E. "A Critical Analysis of Scientific Research in Spelling." Unpublished Ed.D. dissertation, University of Pittsburgh, 1956. 246p. UM 56-3826.

Russell, David H. "Interrelationships of the Language Arts and Personality," *Elementary English,* XXX (March, 1953), 167-180.

Sebeok, Thomas A., ed. *Style in Language.* New York: John Wiley, 1960. 470p.

Shane, Harold G. *Research Helps in Teaching the Language Arts.* Washington: Association for Supervision and Curriculum Development, National Education Association, 1955. 80p.

Smith, Nila B., ed. *Areas of Research Interest in the Language Arts.* Champaign, Illinois: National Council of Teachers of English, 1952. 36p. O.P.

Strom, Ingrid M. "Research in Grammar and Usage and Its Implications for Teaching Writing." *Bulletin of the School of Education,* Indiana University, XXXVI, 5 (September, 1960). 23p.

Indices and Abstracts

In addition to the specialized references listed above, the investigator may wish to consult more general sources which include references on written composition and allied topics. *Education Index* is, of course, an invaluable reference to magazine, journal, and bulletin articles. Far less overlapping with the above sources are the various indices to unpublished studies:

Dissertation Abstracts. Abstracts of all doctoral dissertations completed at the many American and Canadian universities which cooperate with University Microfilms. Dissertations of other universities are also listed, by author and subject, in *Index to American Doctoral Dissertations,* published annually as a supplement to *Doctoral Dissertations.*

Master's Theses in Education. Includes American and Canadian theses.

Research Studies in Education. Differs from the two above sources in that it includes studies in progress as well as studies completed.

Index to Theses Accepted for Higher Degrees in the Universities of Great Britain and Ireland. Includes dissertations and theses.

Bibliography for This Study

Mainly because it may be helpful to investigators seeking a bibliography on the teaching of composition but also because it reveals the scope and limits of the studies screened in preparing this document, a complete list follows of all references examined by the Committee on the State of Knowledge about Composition, cooperating colleagues, and the directors of this project. Some of the items are not research at all; many are reports of projects which might best be described as "pilot" or as "local" studies; many seem incompletely described; many are experiments based entirely on objective tests and hence were not considered sufficiently valid for the purposes of fuller consideration in this report. If an entry is preceded by an asterisk, that study is considered in one of the other chapters in this report. If an entry is followed by "UM" and a number, that is the number to use in ordering a microfilm of the study from University Microfilms, 313 North First Street, Ann Arbor, Michigan. Such studies are not available in any form from the institutions at which they were written. Unpublished dissertations and theses without the "UM" designation either were screened by cooperating colleagues on campuses other than that of the University of Iowa or were brought to Iowa City by interlibrary loan or by microfilm ordered directly from the institution at which the study was conducted.

1. Aaro, Ruth. "Types of Errors Found in Business Letters," *National Business Education Quarterly*, XIV (May, 1946), 23-28.
2. Adams, Robert J. "Experiment in the Teaching of English Composition," *Journal of Experimental Education*, V (June, 1937), 339-342.
3. Allport, Floyd H.; Lynette Walker; and Eleanor Lathers. "Written Composition and Characteristics of Personality," *Archives of Psychology*, No. 173 (December, 1934). 82p.

* 4. Anderson, C. C. "The New STEP Essay Test as a Measure of Composition Ability," *Educational and Psychological Measurement*, XX (Spring, 1960), 95-102.

* 5. Anderson, Edward L. "A Study of Short Stories Written by Students in College Composition Classes to Determine Relationships Between the Prior Experiences of the Students and Their Treatment of Setting and Character." Unpublished Ph.D. dissertation, New York University, 1950. 557p. UM 1919.

6. Anderson, Eleanor M., *et al.* "Construction and Evaluation of a Scale for Creative Writing." Unpublished M.A. thesis, Boston University, 1957.
7. Anderson, Harold A., and Arthur E. Traxler. "The Reliability of the Reading of an English Essay Test—A Second Study," *School Review*, XLVIII (September, 1940), 521-530.

* 8. Anderson, John E. "An Evaluation of Various Indices of Linguistic Development," *Child Development*, VIII (March, 1937), 62-68.

9. Angell, Melvin A. "High School English: Can We Justify It as a College-Admission Requirement?" *School Review*, LXI (September, 1953), 351-356.
10. Angene, Lester. "Proposals for the Conduct of Written Composition Activities in the Secondary School Inherent in an Analysis of the Language Composition Act." Unpublished Ph.D. dissertation, The Ohio State University, 1956. UM 56-1111.
11. Ash, I. O. "An Experimental Evaluation of the Stylistic Approach in Teaching Written Composition in the Junior High School," *Journal of Experimental Education*, IV (September, 1935), 54-62.
12. Asker, William. "Does Knowledge of Formal Grammar Function?" *School and Society*, XVII (January 27, 1923), 109-111.
13. Assisi, Sister Mary Frances. *An Analysis of the Contents of Children's Inventive Compositions.* Washington: Catholic University of America Press, 1950. 28p. Also in *Catholic Educational Review*, XLVIII (September, 1950), 441-454.
14. Avery, Curtis E., and E. G. Williamson. "Achievement of Students in Subfreshman Composition," *Journal of Educational Psychology*, XXIX (April, 1938), 257-267.
15. Averill, Lawrence A. "Some Uses of the ACE English Test in Worcester Teachers College," *School and Society*, LXI (April 21, 1945), 253-255.
16. Ayer, Fred C. "An Evaluation of High School Spelling," *School Review*, LIX (April, 1951), 208-219.

17. Aymes, Alan. "An Evaluation of a Series of Sixth Grade Compositions to Determine Growth in Originality, Elaborative Thinking and Vocabulary." Unpublished Ph.D. dissertation, Boston University, 1959.
18. Babcock, Clarence M. "An Investigation of the Relationships between Occupational Interest and Achievement in Written Composition." Unpublished Ed.D. dissertation, University of Denver, 1948.
19. Bacigalupo, Edwina C. "Evaluation of the Results of a Language Vocabulary Test for Grade 6 in Comparison to Written Composition." Unpublished M.A. thesis, Boston University, 1958.
20. Bagley, Dorothy. "Critical Survey of Objective Estimates in the Teaching of English," *British Journal of Educational Psychology,* VII (February and June, 1937), 57-71 and 138-155.
21. Baker, William D. "An Investigation of Characteristics of Poor Writers," *College Composition and Communication,* V (February, 1954), 23-27.
22. Banerjee, Hemchandra. "Reliability of Repeated Grading of English Essays," *Indian Journal of Psychology,* XIV (January, 1939), 43-50.
23. Barch, A. M., and R. L. Wright. "The Background and Self-Picture of Good and Poor Writers," *Journal of Communication,* VII (Winter, 1957), 192-193.
24. Barlow, Elizabeth R. "Improving Sentence Patterns in Written Composition in Grade Nine." Unpublished Ed.D. dissertation, Boston University, 1954. 188p.
25. Barnes, Walter. "Certain Aspects of the Out-of-School Language Activities of Children of the Seventh, Eighth, and Ninth Grades." Unpublished Ph.D. dissertation, New York University, 1930.
26. Barton, Dorothy J. "The Relation of Personality to Composition Writing and to Style of Writing." Unpublished Ph.D. dissertation, Syracuse University, 1946.
27. Bates, Lorelis P. "Incompetencies in English Found Among Freshmen Students in the State Colleges of Louisiana." Unpublished M.A. thesis, Northwestern State College, 1960.
28. Bear, Mata V. "Children's Growth in the Use of Written Language," *Elementary English Review,* XVI (December, 1939), 312-319.
29. Beatty, Dorothy M. "A Comparison of Two Methods of Teaching Spelling." Unpublished Ph.D. dissertation, University of Iowa, 1955. 320p.
30. Beck, Mildred L. "The Construction and Validation of an Analytical Handwriting Scale for Grades IV, V, and VI." Unpublished Ph.D. dissertation, Boston University, 1956.
• 31. Becker, Samuel L., *et al. Communication Skills: An Experiment in Instructional Methods.* Iowa City: University of Iowa, 1958. 43p. Synopsis and some additional analyses, as "The Effect of Instructional Methods upon Achievement and Attitudes in Communication Skills," by Samuel L. Becker and Carl A. Dallinger, in *Speech Monographs,* XXVII (March, 1960), 70-76.
32. Beckman, Vernon E. "An Investigation and Analysis of the Contributions to Critical Thinking Made by Courses in Argumentation and Discussion in Selected Colleges." Unpublished Ph.D. dissertation, University of Minnesota, 1956. 144p. UM 56-5373.
33. Bellows, Susie M. "Form and Content of Fourth Grade Children's Writings." Unpublished M.A. thesis, George Peabody College for Teachers, 1938.
34. Berko, Jean. "The Child's Learning of English Morphology." Unpublished Ph.D. dissertation, Radcliffe College, 1956.
35. Betz, Annette, and Esther Marshall. "Grammar Based on Errors," *English Journal,* V (September, 1916), 491-500.
36. Bezzi, Diodato, "Standard Manuscript Scales for Grades I, II and III." Unpublished Ed.D. dissertation, University of Oklahoma, 1956. 91p. UM 19,485.
37. Binney, James, "Short Report Concerning Grammar," *Education,* LXXVI (April, 1956), 472-477.

38. Biscoe, William S. "A Study of Children's Language Behavior in Written Composition." Unpublished Ph.D. dissertation, Boston University, 1951.

39. Blair, Mary. "A Study of the Types and Causes of Sentence Structure Errors Made by Typical Junior High School Pupils." Unpublished M.A. thesis, New York University, 1938.

40. Blood, Doris M. "Independent Writing in Grades 1 and 2." Unpublished Ph.D. dissertation, Boston University, 1959.

41. Boardman, Anne C. "A Study of the Proficiency of College Freshmen and Graduating High School Seniors in Certain Elements of English Composition." Unpublished Ph.D. dissertation, University of Minnesota, 1936.

• 42. Boder, David P. "The Adjective-Verb Quotient: A Contribution to the Psychology of Language," *Psychological Record*, III (March, 1940), 310-343.

43. Bontrager, O. Ray. "An Experimental Appraisal of Pupil Control as a Factor in the Grade Placement of Certain Items of Punctuation." Unpublished Ph.D. dissertation, University of Iowa, 1934.

44. Borass, Harold O. "Photographic Analysis of Certain Letter Forms with Respect to Speed Changes and Stability," *Journal of Experimental Education*, XX (September, 1951), 87-96.

45. Bown, Robert F., and Dorothy R. Smith. "Integration of English Usage and Typing Skills," *Clearing House*, XII (December, 1937), 233-237.

46. Boyd, William. "The Development of Sentence Structure in Childhood," *British Journal of Psychology* (General Section), XVII (January, 1927), 181-191.

47. Boyet, Aggie. "The Effect of Student Group Editing of Themes on the Improvement of Writing in Freshman English Composition." Unpublished Ed.D. dissertation, Oklahoma Agricultural and Mechanical College, 1955. 144p.

48. Bradford, Leland P. "Study of Certain Factors Affecting English Usage," *Journal of Educational Research*, XXXV (October, 1941), 109-118.

49. Briggs, T. H. "Formal English Grammar as a Discipline," *Teachers College Record*, XIV (September, 1913), 251-343.

50. Brittain, Frances J., and J. A. Fitzgerald. "The Vocabulary and Spelling Errors of Second Grade Children's Themes," *Elementary English Review*, XIX (February, 1942), 43-50.

51. Britton, Wayne L. "An Analytical Study of the Abilities of a Group of University of Illinois Freshmen in Certain Selected Aspects of the Organization of Ideas and Materials for Communication." Unpublished Ph.D. dissertation, University of Illinois, 1948.

52. Buck, Dorothy S.; Evelyn J. Paxton; and Roger Hyndman. *Report on Reading Assistants for High School English Teachers: 1960-1961*. Los Angeles: Division of Secondary Education, Los Angeles City School Districts, 1961. 17p.

• 53. Buckingham, Burdette R., and E. W. Dolch. *A Combined Word List*. Boston: Ginn and Company, 1936. 185p.

54. Burke, Armand F. "Considerations for Developing the Freshman English Course at Brockport State Teachers College." Unpublished Ed.D. dissertation, Teachers College, Columbia University, 1950.

• 55. Burrows, Alvina T., *et al. They All Want to Write*. Rev. ed. Englewood Cliffs: Prentice-Hall, 1952. 240p.

56. Bushnell, Paul P. *An Analytical Contrast of Oral with Written English*. Contributions to Education, No. 451. New York: Bureau of Publications, Teachers College, Columbia University, 1930. 87p.

• 57. Butterfield, Clair J. "The Effect of a Knowledge of Certain Grammatical Elements on the Acquisition and Retention of Related Punctuation Skills." Unpublished Ph.D. dissertation, University of Iowa, 1945.

• 58. Buxton, Earl W. "An Experiment to Test the Effects of Writing Frequency and Guided Practice upon Students' Skill in Written Expression." Un-

published Ph.D. dissertation, Stanford University, 1958. 135p. UM 58-3596. Synopsis in *Alberta Journal of Educational Research*, V (June, 1959), 91-99.
59. Calhoun, Robert T. "A Comparison of a Typical and an Intensive Method of Teaching Spelling," *Elementary School Journal*, LV (November, 1954), 154-157.
60. Callewaert, H. "A Rational Technique of Handwriting," *Journal of Educational Research*, XLI (September, 1947), 1-12.
61. Carlson, Ruth K. "Stimulating Children in Grades Four, Five, and Six to Write Original Stories." Unpublished Ph.D. dissertation, University of California, Berkeley, September, 1960.
62. Carpenter, William, *et al.* "The Use of Closed Circuit Television at Evanston Township High School." Evanston, Illinois: Evanston Township High School, [n.d.]. Mimeographed.
63. Carr, William L. "By Their Fruits," *Classical Journal*, XXXVII (March, 1942), 334-350.
64. Carroll, John. "A Factoral Analysis of Verbal Abilities." Unpublished Ph.D. dissertation, University of Minnesota, 1941.
65. Carsley, J. D. "The Guidance of Extensive Reading at the Top of the Secondary Modern School," *Educational Review*, VII (June, 1955), 185-198.
* 66. Cast, B. M. D. "The Efficiency of Different Methods of Marking English Composition," *British Journal of Educational Psychology*. IX (November, 1939), 257-269, and X (February, 1940), 49-60.
67. Castelnuovo-Tedesco, Peter. "A Study of the Relationship between Handwriting and Personality Variables," *Genetic Psychology Monographs*, XXXVII (May, 1948), 167-220.
68. Cawley, F. "The Difficulty of English Grammar for Pupils of Secondary School Age." Unpublished M.Ed. thesis, Manchester University, 1957. 89p.
69. Chalifour, Jessie B. "The Effect of Training Students to Grade Compositions on Composition Work." Unpublished M.S. thesis, Butler University, 1937.
70. Chiles, Mildred H. "Social Relationships as a Stimulus for Creative Writing in the Elementary School: Concepts toward Debt." Unpublished M.Ed. thesis, University of Texas, 1959. 162p.
71. Chisholm, Francis P. "Grammatical Structure in Linguistic Behavior: An Introduction to the General Study of Language." Unpublished Ph.D. dissertation, Syracuse University, 1944.
* 72. Chotlos, John W. "A Statistical and Comparative Analysis of Individual Written Language Samples." Ph.D. dissertation, University of Iowa, 1942. Published in *Psychological Monographs*, LVI (1944), 77-110.
73. Christophe, Maurice M. "A Study of English Proficiency of Freshman Students at Arkansas Baptist College." Unpublished M.A. thesis, Alabama State College, 1960.
74. Clark, Gwyn R. "How Well Do Sixth Graders Write?" *Peabody Journal of Education*, XXXII (July, 1954), 33-42.
75. Clark, Mrs. Ray W. "An Experimental Study Comparing Two Approaches to Written Composition in the Eighth Grade." Unpublished M.Ed. thesis, University of Texas, 1958.
76. Clark, Willis W. "A Study of Factors Related to Mastery of Skills in Reading, Arithmetic, and Written Expression in Los Angeles Senior High School." Unpublished Ph.D. dissertation, University of Southern California, 1941.
77. Clegg, William C. "The Effectiveness of Practice in Simulated Situations as Training for Actual Use of a Skill." Unpublished Ph.D. dissertation, University of Toronto, 1957.
78. Coale, Willis B., and M. E. Smith. *Successful Practices in the Teaching of English to Bilingual Children in Hawaii.* U. S. Department of Interior, Office of Education, Bulletin No. 14. Washington: Government Printing Office, 1937. 163p.
79. Cochrane, James E. "An Analysis of the Transition from High School to College in English Composition in Selected New York State Schools." Unpublished Ph.D. dissertation, University of Connecticut, [n.d.].

*80. Coleman, J. H. *Written Composition Interests of Junior and Senior High School Pupils.* Contributions to Education, No. 494. New York: Bureau of Publications, Teachers College, Columbia University, 1931. 117p.

81. Colvin, S. S. "Invention Versus Form in English Composition: An Inductive Study," *The Pedagogical Seminary,* IX (December, 1902), 393-421.

82. ——————, and I. F. Meyer. "Imaginative Elements in the Written Work of School Children," *The Pedagogical Seminary,* XIII (March, 1906), 84-93.

83. Cook, Luella. "Teaching Grammar and Usage in Relation to Speech and Writing," *Elementary English Review,* XXIII (May, 1946), 193-198.

84. Cooper, Elizabeth K. "A Study of Children's Creative Writing." Unpublished Ed.D. dissertation, University of California, Los Angeles, 1947. 230p.

85. Coward, Ann F. "A Comparison of Two Methods of Grading English Compositions," *Journal of Educational Research,* XLVI (October, 1952), 81-93.

86. Cox, Homer L. "Factors Involved in Clarity of Language and Evaluation of Clarity Achieved in Selected Business Letters." Unpublished Ph.D. dissertation, Northwestern University, 1955.

87. Cromwell, Otelia. "Preparation for Freshman Composition," *English Journal,* XXV (September, 1936), 551-556.

*88. Culpepper-Hagen, Lessie. "An Investigation of General Factors Relating to the Writing Effectiveness of Freshmen at the University of Denver." Unpublished Ed.D. dissertation, University of Denver, 1950. 154p. + unpaginated appendices.

89. Cunningham, Mary R. "A Demonstration of the Effectiveness of Certain Remedial Exercises in Correcting Faulty Sentence Structure." Unpublished M.A. thesis, Rhode Island College of Education, 1957.

90. Currie, Caroline. "The Relationship of Certain Selected Factors to Achievement in Freshman Composition." Unpublished Ph.D. dissertation, Northwestern University, 1957. 167p. UM 24,901.

91. Cutright, Prudence. "A Comparison of Methods of Securing Correct Language Usage," *Elementary School Journal,* XXXIV (May, 1934), 681-690.

92. Daniel, J. T., and George L. Hinds. "An Introductory Study of Identification Reactions in Reading, Writing, Speaking, Listening." Unpublished M.A. thesis, University of Denver, 1946.

93. Davis, Edith A. *The Development of Linguistic Skill in Twins, Singletons with Siblings, and Only Children from Age Five to Ten Years.* Institute of Child Welfare Monograph Series, No. XIV. Minneapolis: University of Minnesota Press, 1937.

94. ——————. "The Location of the Subordinate Clause in Oral and Written Language," *Child Development,* XII (December, 1941), 333-338.

95. Davis, Lavonne B. "A Study of the Vocabulary Used in Sixth Grade Compositions." Unpublished M.A. thesis, San Diego State College, 1960.

96. Davison, Hugo M. "Factors in the Scholastic Ability of College Freshmen." Unpublished Ph.D. dissertation, Harvard University, 1950.

97. Day, Margaret E. "An Analysis of the Factors Making for Success or Failure in Freshman Writing." Unpublished M.A. thesis, University of Minnesota, 1944. Also see "How Well Do High School Graduates Write?" *English Journal,* XXXII (November, 1943), 493-499.

*98. Deacon, Lawrence S. "A Comparison of Two Programs of Instruction in Spelling." Unpublished Ed.D. dissertation, University of California, Berkeley, 1955. 118p.

99. Dell, Homer W. "A Study of the Errors of Wood Township Grade and High School Pupils in Their Use of the Mechanics of Written English Expression." Unpublished M.S. thesis, Pennsylvania State College, 1937. 37p.

100. Derobbio, Louis. "Research and Lesson Plans on Children's Writing." Unpublished Ph.D. dissertation, Boston University, [n.d.].

*101. Diederich, Paul B.; John W. French; and Sydell T. Carlton. *Factors in Judgments of Writing Ability.* Research Bulletin RB-61-15. Princeton: Educational Testing Service, 1961. 60p. + 31 pp. of appendices.

3

*102. ————. "The Measurement of Skill in Writing," *School Review*, LIV (December, 1946), 584-592.

103. ————. "Teaching English with Test Exercises," *School Review*, LV (February, 1947), 80-86.

104. ————. "Testing in the New English Program," *English Record*, III (Spring, 1953), 11-17.

105. Dolch, Edward W. "Tested Word Knowledge vs. Frequency Counts," *Journal of Educational Research*, XLIV (February, 1951), 457-470.

106. Dollard, John, and O. Hobart Mowrer. "A Method of Measuring Tension in Written Documents," *Journal of Abnormal and Social Psychology*, XLII (January, 1947), 3-32.

107. Donaldson, Robert G. "Implications of Linguistics for the Teaching of English Composition." Unpublished M.Ed. thesis, Wayne State University, 1955.

108. Douglass, Dorothy M. "An Analysis of the Frequency and Nature of Misspellings in the Class Work of 128 Fifth Grade Children." Unpublished M.A. thesis, Boston University, 1956.

109. Douglass, Harl R., and Clifford Kittelson. "The Transfer of Training in High School Latin to English Grammar, Spelling, and Vocabulary," *Journal of Experimental Education*, IV (September, 1935), 26-33.

110. Dressel, Paul; John Schmid; and Gerald Kincaid. "The Effect of Writing Frequency Upon Essay-Type Writing Proficiency at the College Level," *Journal of Educational Research*, XLVI (December, 1952), 283-293.

111. Durkee, Frank M. "A Study of Written Composition in the Freshman Course in English in the Colleges of New Jersey." Unpublished Ph.D. dissertation, Rutgers University, 1951.

*112. Dusel, William J. "Determining an Efficient Teaching Load in English," *Illinois English Bulletin*, XLIII (October, 1955). 1-19.

113. Eaton, Merrill T. "An Analysis of Factors Related to the Language Arts Achievement of Sixth-Grade Pupils." Bulletin No. 8. Indianapolis, Indiana: Department of Public Instruction, 1944.

114. Eberhart, Wilfred. "The Teaching of Functional Grammar in the Secondary School." Unpublished Ph.D. dissertation, The Ohio State University, 1936.

115. Edmiston, R. W., and C. N. Gingerich. "The Relation of Factors of English Usage to Composition," *Journal of Educational Research*, XXXVI (December, 1942), 269-271.

*116. Edmund, Neal R. "A Study of the Relationship Between Prior Experiences and the Quality of Creative Writing by Seventh Grade Pupils." Unpublished Ed.D. dissertation, Syracuse University, 1956. 117p. UM 18.016. Synopses in *Journal of Educational Research*, LI (March, 1958), 481-492, and in *Elementary English*, XXXV (April, 1958), 248-249.

*117. Eley, Earle G. "An Analysis of Writing Competence." Unpublished Ph.D. dissertation, University of Chicago, 1953. 108p.

*118. Emery, Donald W. *Variant Spellings in Modern American Dictionaries.* Champaign, Illinois: National Council of Teachers of English, 1958. 43p.

*119. Enochs, James B. "Measuring the Ability to Write." Unpublished Ph.D. dissertation, University of Chicago, 1948. 104p.

120. Enstrom, Eric A. "The Extent of the Use of the Left Hand in Handwriting and the Determination of the Relative Efficiency of the Various Hand-Wrist-Arm-Paper Adjustments." Unpublished Ph.D. dissertation, University of Pittsburgh, 1957. 167p. UM 21,001.

*121. Erickson, Lawrence W., and Cletus Clow. *The Manual Portable Typewriter as a Tool of Learning with Fifth-Grade Elementary School Pupils.* New York: Teachers College, Columbia University, 1959.

122. Evans, E. G. S. "An Experimental Study of Methods of Teaching English," *British Journal of Educational Psychology*, XXIII (June, 1953), 127-128.

*123. Evans, James W. "The Social Importance and the Pupil Control of Certain Punctuation Variants." Unpublished Ph.D. dissertation, University of Iowa, 1939. 232p.

124. Ewan, Stacy N., Jr. "The Relation of Class Size and Teaching Method to Pupil Achievement." Unpublished Ph.D. dissertation, University of Pensylvania, 1935.

125. Eysenck, Hans J. "Graphological Analysis and Psychiatry: An Experimental Study," *British Journal of Psychology*, XXXV (May, 1945), 70-81.

126. Fea, Henry R. "Some Relationships among Abilities in Oral Language, Written Language, and Reading of Intermediate Grade Children." Unpublished Ph.D. dissertation, University of California, Berkeley, 1950. 174p. Synopsis in *Journal of Educational Psychology*, XLIV (March, 1953), 159-175.

127. Fee, Mary. "Functional Grammar and Its Relation to Correct English Usage." Unpublished Ph.D. dissertation, University of Kansas, 1940. 121p.

128. Fellers, Alvin L. "Problems in Writing in College Composition Classes." Unpublished Ed.D. dissertation, Stanford University, 1953. 365p. UM 6868.

*129. Fellows, John E. "The Influence of Theme-reading and Theme-correction on Eliminating Technical Errors in the Written Compositions of Ninth Grade Pupils." University of Iowa *Studies in Education*, VII, 1 (1932).

*130. Finlayson, Douglas S. "The Reliability of the Marking of Essays," *British Journal of Educational Psychology*, XXI (June, 1951), 126-134.

*131. Fitzgerald, James A. *A Basic Life Spelling Vocabulary*. Milwaukee: Bruce Publishing Company, 1951. 161p.

132. ————. "The Letter Writing Difficulties of Intermediate Grade Children," pp. 332-338 in *Language Arts in the Elementary School*. Twentieth Yearbook of the Department of Elementary School Principals. Washington: National Education Association, 1941.

135. ————. "The Vocabulary, Spelling Errors, and Situations of Children's Letters Written in Life Outside School." University of Iowa *Studies in Education*, IX, 1 (1934).

134. ————, and Lawrence A. Knaphle. "Crucial Language Difficulties in Letter Writing of Elementary-School Children," *Elementary English Review*, XXI (January, 1944), 14-19.

135. Flesch, Rudolf. "Measuring the Level of Abstraction," *Journal of Applied Psychology*, XXXIV (December, 1950), 384-390. Also see James J. Jenkins and Robert L. Jones, "Flesch's 'Measuring the Level of Abstraction'," and Flesch's "Reply to Criticism by Jenkins and Jones," *Journal of Applied Psychology*, XXXV (February, 1951), 68-69.

136. Flynn, Helen, and Stephen M. Corey. "Teaching Communication through the Use of Sound Films," *School Review*, LIII (June, 1945), 348-352.

137. Fogler, Sigmund. "The Case for a Basic Written Vocabulary," *Elementary School Journal*, XLVII (September, 1946), 43-47.

138. Ford, Clarence T. "Developments in Written Composition During the Primary School Period," *British Journal of Educational Psychology*, XXIV (February, 1954), 38-45.

139. Forester, John J. "Differences between Typed and Handwritten Elementary School Narrative Compositions." Unpublished Ph.D. dissertation, New York University, 1934.

140. Foster, Harry K. "The Semantic Variations of Certain High Frequency Words in the Written Compositions of Eighth Grade Pupils." Unpublished Ph.D. dissertation, University of Iowa, 1940.

141. Foster, Richard R. "Some Aspects of Teaching and Testing in College Freshman English." Unpublished Ph.D. dissertation, University of Iowa, 1928.

142. Fox, William H. "Spelling Proficiency in Township Schools in Indiana," *Bulletin of the School of Education*, Indiana University, XXIII, 1 (January, 1947).

143. ————, and Merrill T. Eaton. "Analysis of the Spelling Proficiency of 82,833 Pupils in Grades 2 to 8 in 3547 Teaching Units of the City Schools in Indiana." *Bulletin of the School of Education*, Indiana University, XXII, 2 (March, 1946).

144. Freeman, Frank N. "Survey of Manuscript Writing in the Public Schools," *Elementary School Journal*, XLVI (March, 1946), 375-380.

145. ————. *Teaching Handwriting: What Research Says to the Teacher.* Bulletin No. 4, Department of Classroom Teachers and American Educational Research Association. Washington: National Education Association, 1954.

146. French, John W. "School of Thought in Judging Excellence of English Themes," pp. 19-28 in *Proceedings of the 1961 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 1962.

•147. ————. "What English Teachers Think of Essay Testing," *English Journal,* XLVI (April, 1957), 196-201.

148. Fries, Charles C. *American English Grammar.* National Council of Teachers of English Monograph No. 10. New York: Appleton-Century-Crofts, 1940. 313p.

149. Frogner, Ellen. "Problems of Sentence Structure in Pupils' Themes," *English Journal,* XXII (November, 1933), 742-749.

150. ————. "A Study of the Relative Efficacy of Grammatical and a Thought Approach to the Improvement of Sentence Structure in Grades Nine and Eleven." Unpublished Ph.D. dissertation, University of Minnesota, 1938. 681p. Synopses in *English Journal,* XXVIII (September, 1939), 518-526, and in *School Review,* XLVII (November, 1939), 663-675.

151. Fulbright, John B. "An Experimental Study of the Relative Effectiveness of the Use of Pictures for Teaching Selected Basic Language Skills at the Sixth Grade Level as Compared with the Effectiveness of the Procedures Outlined in the Textbook Manual." Unpublished Ed.D. dissertation, Colorado State College of Education, 1957.

152. Fuller, Max E. "A Study of the Effectiveness of a Course in Basic Communication." Unpublished Ph.D. dissertation, Northwestern University, 1950.

153. Gaiser, Gerhard W. "Improving Techniques in the Prose Composition of Juniors and Seniors in High School." Unpublished M.A. thesis, The Ohio State University, 1937. 155p.

154. Garnett, Wilma L. "A Study of Status and Improvement of College Freshmen in Certain Skills of English Composition." Unpublished Ph.D. dissertation, University of Iowa, 1937. Synopsis in *Journal of Experimental Education,* VI (September, 1937), 29-34.

155. Gaskill, Harold Vincent. "Language Responses and Intelligence; Verbalization and Intelligence," *Journal of Genetic Psychology,* LVIII (June, 1941), 407-417.

156. Geist, Robert J. "Agronomy-Rhetoric Experiment," *School Science and Mathematics,* IL (January, 1949), 1-5.

157. Geoghegan, Patricia S., and James A. Fitzgerald. "Composition Errors in Letters Written by Children Outside the School," *Elementary School Journal,* XXXV (June, 1935), 768-775.

•158. Gerber, John C. "Testing and Evaluation in the Skills of Communication," *College English,* IX (April, 1948), 375-384.

159. Giedt, Eugene E. "An Experimental Study of Two Methods of Teaching Composition to High School Sophomores." Unpublished M.A. thesis, Sacramento State College, 1957.

•160. Gillie, Paul J. "A Simplified Formula for Measuring Abstraction in Writing," *Journal of Applied Psychology,* XLI (August, 1957), 214-217. Also see Haskins, *op. cit.* below.

161. Goodman, J. H. "Growth in Punctuation and Capitalization Abilities," *Journal of Educational Research,* XVIII (November, 1934), 195-202.

162. Goodner, Mildred. "An Analysis of the Spelling Errors of College Freshmen," *Illinois English Bulletin,* XXXV (May, 1948), 9-17.

163. Gordon, Hans C. "Composition Achievement in Reorganized Plan of Teaching English in Senior High School." *Practical Values of Educational Research: Official Report.* American Educational Research Association. Washington: National Education Association, 1938.

*164. Goyer, Robert S. "A Study of Individual Differences in Ability and Achievement of College Students in the Organization of Ideas." Unpublished Ph.D. dissertation, The Ohio State University, 1955. 113p. UM 56-1604.

165. Gramlich, Francis W. "Symbolism and Meaning." Unpublished Ph.D. dissertation, Princeton University, 1936.

166. Grant, Albert, and Margaret M. Marble. "Results of Cincinnati Handwriting Survey," *School Review*, XLVIII (November, 1940), 693-696.

167. Gray, Mary E. "An Analysis of Language Themes in Grade Five, Grade Eight, and Grade Eleven." Unpublished M.Ed. thesis, University of Alberta, 1959.

168. Green, James R. "A Comparison of Oral and Written Language." Unpublished Ph.D. dissertation, New York University, 1958. 215p. UM 58-7618.

169. Greene, Harry A. *A Criterion for a Course of Study in Mechanics of Composition.* University of Iowa *Studies in Education,* Monograph No. 4. Iowa City: University of Iowa, 1933.

170. ———. "Direct Versus Formal Methods in Elementary English," *Elementary English,* XXIV (May, 1947), 273-285.

171. Grego, Richard F. "The Communication Activities of 120 Graduates of the New York City Community College." Unpublished Ed.D. dissertation, Columbia University, 1957.

172. Grise, Presley M. "A Study of the Relationships of Written Composition Ability of High-School Graduates." Unpublished Ph.D. dissertation, University of Kentucky, 1940. 290p.

173. Grundy, Ernest B. "Applying Language Principles to Written Composition." Unpublished M.A. thesis, Arizona State College, 1958.

174. Guiler, Walter S. "Analysis of Capitalization Errors," *English Journal,* XX (January, 1931), 21-26.

175. ———. "Capitalization Disabilities of College Freshmen," *Journal of American Association of Collegiate Registrars,* XXII (April, 1947), 317-327.

176. ———. "Diagnosing Student Shortcomings in English Composition," *Journal of Educational Research*, XIV (September, 1926), 112-119.

177. ———. "Disabilities of College Freshmen in Sentence Structure," *School Review,* LIV (October, 1946), 480-487.

178. ———. "Grammatical-Usage Disabilities of College Freshmen," *Journal of American Association of Collegiate Registrars,* XXII (October, 1946), 49-57.

179. ———. "Punctuation Disabilities of College Freshmen," *Journal of American Association of Collegiate Registrars,* XXII (January, 1947), 183-191.

180. ———. "Spelling at the College Level." *Journal of the American Association of Collegiate Registrars,* XX (October, 1944), 96-105.

181. Gunn, M. Agnella. "A Technique for Improving Basic Skills in English in High School." University of Iowa *Studies in Education,* VIII, 7 (1934).

182. Hale, R. L. "Composition and Observation," *The School* (Secondary Edition), XXVIII (January, 1940), 413-416.

183. Halvorsen, Gladys C. "Some Effects of Emphasizing Mechanical Accuracy on Written Expression of Sixth-Grade Children." Unpublished Ph.D. dissertation, University of California. Berkeley, 1960.

*184. Halverson, Nelius O. "Two Methods of Indicating Errors in Themes," *College English,* II (December, 1940), 277-279.

185. Hamlin, Roy M. "An Analysis of Age-Progress Curves as Related to Mental Growth Curves: A Study of Vocabulary." Unpublished Ph.D. dissertation, Columbia University, 1945.

186. Harap, Henry. "The Most Common Grammatical Errors," *English Journal,* XIX (June, 1930), 440-446.

*187. Harrell, Lester E., Jr. "An Inter-Comparison of the Quality and Rate of the Development of Oral and Written Language in Children." Ph.D. dissertation, University of Minnesota, 1956. 204p. UM 56-1732. Also, as "A Comparison of the Development of Oral and Written Language in

School-Age Children," in *Monographs of the Society for Research in Child Development,* XXII, 3 (1957). 77p.

188. Harris, James H. "An Inquiry into the Compositional Interests of Pupils in the Seventh or Eighth Grades," *English Journal,* II (January, 1913), 34-43.

189. Harris, Oliver E. "An Investigation of Spelling Achievement of Secondary School Pupils," *Educational Administration and Supervision,* XXXIV (April, 1948), 208-219.

*190. Harris, Roland J. "An Experimental Inquiry into the Functions and Value of Formal Grammar in the Teaching of English, with Special Reference to the Teaching of Correct Written English to Children Aged Twelve to Fourteen." Unpublished Ph.D. dissertation, University of London, 1962. 291p.

191. Harston, L. D. "A Five Year Study of Objective Tests for Sectioning Courses in English Composition," *Journal of Applied Psychology,* XIV (June, 1930), 202-210.

192. Hartog, Sir Philip, *et al. The Marking of English Essays.* London: Macmillan Company, 1941.

*193. ————, and E. C. Rhodes. *The Marks of Examiners.* London: Macmillan Company, 1941.

*194. Hartsig, Barbara A. "A Study of the Creative Writing of Children in the Intermediate Grades." Unpublished Ed.D. dissertation, University of California, Los Angeles, 1956. 389p.

*195. Haskins, Jack B. "Validation of the Abstraction Index as a Tool for Content-Effects Analysis and Content Analysis," *Journal of Applied Psychology,* XLIV (April, 1960), 102-106.

196. Hazelrigg, Blanche. "An Experiment on Teaching Composition," *English Journal,* XXII (June, 1933), 486-490.

197. Healy, Katharine H. "Study of the Factors Involved in the Rating of Pupils' Compositions," *Journal of Experimental Education,* IV (September, 1935), 50-53.

198. Heed, Helen. "Children Certainly Do Improve Their Writing Skills," *Clearing House,* XXX (February, 1956), 365-368.

199. Heider, Fritz, and Grace M. Heider. "A Comparison of Sentence Structure of Deaf and Hearing Children," *Psychological Monographs,* LII, 232 (1940), 42-103. Also in *Volta Review,* XLIII (June, September, and October, 1941), 364-367, 406, 536-540, 599-604, 628, 630.

200. Herber, Harold L. "A Comparison of the Composition Interests of Six Hundred Sixty Students in Grades Eight through Twelve." Unpublished M.A. thesis, Boston University, 1958.

201. Herminghaus, Earl G. "Large-Group Instruction by Television: An Experiment," *School Review,* LXV (Summer, 1957), 119-133.

202. Herrick, Virgil E. "Perception of Handwriting by Bright, Normal, and Dull Children." Unpublished M.A. thesis, University of Wisconsin, 1961.

203. Hewitt, Betty R. "A Survey of Student Deficiencies in Required English Courses in Five New Hampshire Colleges." Unpublished M.A. thesis, Plymouth Teachers College, 1960.

*204. Heys, Frank, Jr. "Report of Experiment in Teaching Composition." October 15, 1961. Mimeographed. 10p. Synopsis, as "The Theme-a-Week Assumption: A Report of an Experiment," *English Journal,* LI, (May, 1962), 320-322. Report of preliminary experiment also in *English Leaflet,* LIX (Fall, 1960), 28-31.

205. Hildreth, Gertrude. "Developmental Problems Associated with Handedness" and "Training of Handedness," *Journal of Genetic Psychology,* LXXVI (March, 1950), 39-100 and 101-144.

206. Hinton, Eugene M. "An Analytical Study of the Qualities of Style and Rhetoric Found in English Composition." Contributions to Education, No. 806. New York: Bureau of Publications, Teachers College, Columbia University, 1940.

207. Hooper, Cyrus L. "The Influence of the Study of Latin on the Student's Knowledge of English Grammar," *English Journal*, I (September, 1912), 393-404.

208. Hoppes, William C. "The Development of Written Expression Among Children of Elementary-School Grades." Unpublished Ph.D. dissertation, University of Chicago, 1931. 144p. Synopsis, as "Considerations in the Development of Children's Language," *Elementary English Review*, XI (March, 1934), 66-70.

•209. Horn, Ernest. *A Basic Writing Vocabulary*. University of Iowa *Monographs in Education*. Iowa City: University of Iowa, 1926. 225p.

210. Horner, Warren B. "The Economy of the Laboratory Method," *English Journal*, XVIII (March, 1929), 214-221.

211. Howell, Miriam M. "The Control of Mechanics of Expression and the Quality of Preceding Experiences as Differentiating Factors in Certain Aspects of Compositions of Seven-Year-Olds." Unpublished Ph.D. dissertation, University of Wisconsin, 1955. 143p.

212. __________. "Spelling through Written Expression," *Elementary School Journal*, LII (December, 1951), 207-214.

213. Howes, Raymond F. "The Talked and the Written," *Quarterly Journal of Speech*, XXVI (October, 1940), 229-235.

•214. Huddleston, Edith M. "Measurement of Writing Ability at the College Entrance Level: Objective vs. Subjective Testing Techniques." Unpublished Ph.D. dissertation, New York University, 1952. Synopsis in *Journal of Experimental Education*, XXII (March, 1954), 165-213.

215. Humphreys, Philo. "A Greater Independence in Written Expression," *Elementary English Review*, XIX (March, 1942), 93-98.

216. Huxtable, Zelma O. "Criteria for Judging Thought Content in Written English," *Journal of Educational Research*, XIX (March, 1929), 188-195.

217. Hwang, Pu. *Errors and Improvement in Rating English Compositions by Means of a Composition Scale*. Contributions to Education, No. 417. New York: Bureau of Publications, Teachers College, Columbia University, 1930. 67p.

218. Irish, Elizabeth H. "The Principles of Time and Motion Applied to the Teaching of Handwriting." Unpublished Ed.D. dissertation, Stanford University, 1948. 133p.

219. Isidro, Antonio. "The Development of Written English Expression of Filipino Children." Unpublished Ph.D. dissertation, University of Chicago, 1935.

220. Jencke, Grace E. *A Study of Précis Writing as a Composition Technique*. Contributions to Education, No. 644. New York: Bureau of Publications, Teachers College, Columbia University, 1935.

221. Jenkins, Russell L. "The Relative Effectiveness of Two Methods of Teaching Written and Spoken English." Unpublished Ed.D. dissertation, Michigan State University, 1952.

222. Jensen, Barry T., and J. W. Jensen. "Shorthand Study and Spelling Abilities," *Journal of Educational Psychology*, XLVI (February, 1955), 112-116.

223. Jersild, Arthur T., and Ruth Ritzman. "Aspects of Language Development: The Growth of Loquacity and Vocabulary," *Child Development*, IX (September, 1938), 243-259.

224. Jewell, Ross M. "Final Report of the Communication Experiment Conducted by the Department of Languages, Speech, and Literature of the Iowa State Teachers College, 1955-1958." Mimeographed. 29p.

225. Johnson, Burges, and Helene Hartley. *An Outline of an Inquiry Being Made at Syracuse University into the Methods, Purposes and Effectiveness of the Training of College Freshmen in Written Composition*. Syracuse: Syracuse University, 1933. 42p.

226. __________, and __________. *Written Composition in American Colleges*. Schenectady, New York: Union College, 1936. 47p.

227. Johnson, Leslie W. "One Hundred Words Most Often Misspelled by Children in the Elementary Grades," *Journal of Educational Research,* LXIV (October, 1950), 154-155.

228. Johnson, W. G., and Ernest G. Mathews. "Errors Most Frequently Checked in Early Freshman Compositions," *Illinois English Bulletin,* XXXI (May, 1944), 1-8, and XXXV (April, 1948), 9-15.

229. Jolley, Nadene E. "Techniques for Improving the Teaching of Written Composition in the Self-Contained Classroom." Unpublished M.A. thesis, The Ohio State University, 1961.

*230. Kaldegg, Gustav. "Substance Symbolism: A Study in Language Psychology," *Journal of Experimental Education,* XVIII (June, 1950), 331-342.

231. Kanfer, Frederick H. "The Effect of Partial Reinforcement on Acquisition and Extinction of a Class of Verbal Responses." Unpublished Ph.D. dissertation, Indiana University, 1953.

232. Karp, Mark. "An Individual Method and a Group Method of Teaching College Freshmen the Mechanics of English Composition," *Journal of Experimental Education,* XI (September, 1942), 9-15.

233. Kaulfers, Walter V. *Four Studies in Teaching Grammar from the Socio-Psychological Viewpoint.* Stanford: Stanford University Press, 1945. 47p.

234. Kaump, Ethel A. "An Analysis of the Structural Differences between the Oral and Written Language of 100 Secondary School Students." Unpublished Ph.D. dissertation, University of Wisconsin, 1940.

235. Kemp, Mabel D. "An Analysis of Creative Writing in a Second Grade Classroom." Unpublished M.S. Thesis, Illinois State Normal University, 1958.

236. Kennedy, Ruth A. M. "A Comparative Study of Two Methods of Teaching Language Arts in the Fifth and Sixth Grades in Douglas Elementary School, Tyler." Unpublished M.A. thesis, Texas College, Tyler, 1959.

237. Kenny, Anna W. "Instruction in Remedial English: A Suggested Program Based on an Analysis of Certain English-Usage Difficulties Among Students in Chicago Public High Schools and Colleges." Unpublished Ph.D. dissertation, University of Chicago, 1945.

238. Khater, Mahmoud R. "The Influence of Social Class on the Language Patterns of Kindergarten Children." Unpublished Ph.D. dissertation, University of Chicago, 1951.

239. Kidd, Elnora L. "A Digest of Approaches to Creative Writing with Primary Children," *Elementary English,* XXV (January, 1948), 47-53.

240. Kiefer, Frieda A., and Paul V. Sangren. "An Experimental Investigation of the Causes of Poor Spelling among University Students, with Suggestions for Improvement," *Journal of Educational Psychology,* XVI (January, 1925), 38-47.

*241. Kimoto, Blanche. "A Quantitative Analysis of Grammatical Categories as a Measure of Dominance." Unpublished Ph.D. dissertation, Boston University, 1951.

*242. Kincaid, Gerald L. "Some Factors Affecting Variations in the Quality of Students' Writing." Unpublished Ed.D. dissertation, (Michigan State College) Michigan State University, 1953. 116p. UM 5922.

243. King, Irvill C. "A Study of the Efficiency of the Method Used by Kansas High School Teachers in Marking Written Compositions." Unpublished M. A. thesis, University of Kansas, 1957.

244. Kingston, Johns, Jr. "A Study of the Relationships Between Students' Grades in College Freshman Composition and Certain Aspects of Their High School English Experience." Unpublished Ph.D. dissertation, Cornell University, 1951.

*245. Kitzhaber, Albert R. *Themes, Theories, and Therapy: The Teaching of Writing in College.* New York: McGraw-Hill Book Company, 1963.

246. Klapper, Hope L. *Closed-Circuit TV as a Medium of Instruction at New York University: 1956-1957.* New York: New York University, 1958.

247. Klineberg, Otto, and Joseph T. Klapper. *The Mass Media: Their Impact on Children and Family Life*. New York: Television Information Office, 1960. Also in *Child Study*, XXXVII (Summer, 1960), 24-32.

248. Knickerbocker, Kenneth L. "Placement of Freshmen in First-Quarter English," *Journal of Higher Education*, XXII (April, 1951), 211-215, 228.

249. Kokhba, Moshelh. "Interest-Patterns in Compositions of Fifth Grade Pupils in American and Palestinian Elementary Schools. Unpublished Ph.D. dissertation, Columbia University, 1937.

250. Kraus, Silvy A. "A Comparison of Three Methods of Teaching Sentence Structure." Unpublished Ph.D. dissertation, University of Minnesota, 1956. 108p. plus 198pp. of unpaginated appendices. UM 20,556. Synopsis in *English Journal*, XLVI (May, 1957), 275-281.

251. Kreitlow, Burton W. "A Study of the Teaching of Letter Writing in the Sixth Grade of Graded and Ungraded Rural Schools." Unpublished Ph.D. dissertation, University of Minnesota, 1950. Synopsis in *Journal of Educational Research*, XLIV (September, 1950), 22-29.

252. Kremer, Josephine. "An Analysis of Sixth-Grade Children's Letters." Unpublished M.A. thesis, University of Iowa, 1931. 42p.

253. Kyte, George C., and Virginia M. Neal. "A Core Vocabulary of Spelling Words," *Elementary School Journal*, LIV (September, 1953), 29-34.

*254. LaBrant, Lou L. "A Study of Certain Language Developments of Children in Grades Four to Twelve, Inclusive," *Genetic Psychology Monographs*, XIV (November, 1933), 387-491.

255. Landmark, Nora. "A Study of the Simple and Expanded Forms of the Verb in the Present Tense." Unpublished Ph.D. dissertation, Michigan State University, 1954.

256. Lange, Phil C. "A Sampling of Composition Errors of College Freshmen in a Course Other Than English," *Journal of Educational Research*, XLII (November, 1948), 191-200.

257. Lannholm, Gerald V. "The Measurement of Ability in Capitalization and Punctuation." Unpublished Ph.D. dissertation, University of Iowa, 1939.

258. Larson, Harold G. "A Comparative Analysis of Errors in Grammar Found in Oral and Written Discourse of First Year Students at University of South Dakota." Unpublished M.A. Thesis, University of South Dakota, 1948.

259. Lauroesch, William. "Improving Preparation for College Writing: A Survey of College Students and College English Instructors to Determine What to Emphasize in College Preparatory Writing." Unpublished Ed.D. dissertation, New York University, 1956. 265p. UM 56-3278.

260. Lawley, D. N., and A. E. G. Pilliner. "Comments on 'Shorter and Longer Essays' by D. M. E. Penfold," *British Journal of Educational Psychology*, XXVII (June, 1957), 142-144.

261. Laycock, Frank. "Spelling Ability and Vocabulary of One Hundred College Freshmen," *Journal of Educational Psychology*, XLV (December, 1954), 485-491.

262. Lazarus, Arnold L. "Grouping Based on High Interest vs. General Ability: A Senior High School Teacher's Viewpoint," *California Journal of Secondary Education*, XXX (January, 1955), 38-41.

263. __________. "Pupils' T.V. Habits," *Educational Leadership*, XIII (January, 1956), 241-242.

264. Lee, Irving. "Emotional Appeal in Rhetorical Theory with Special Reference to Invention, Arrangement, and Style." Unpublished Ph.D. dissertation, Northwestern University, 1939.

265. Leggitt, Dorothy. "Comparison of Abilities in Cursive and Manuscript Writing and in Creative Art," *School Review*, IL (January, 1941), 48-56.

266. Lehrer, Adrienne. "Applications of Linguistics to English in the Secondary School." Unpublished M.A.T. thesis, Brown University, 1960.

267. Lemon, Babette K., and G. T. Buswell. "Oral and Written Expression in Grade IX," *School Review*, LI (November, 1943), 544-549.

268. Lenneberg, Eric H. "An Empirical Investigation into the Relationship between Language and Cognition." Unpublished Ph.D. dissertation, Harvard University, 1956.

*269. Leonard, Sterling A. *Current English Usage*. Chicago: Inland Press, 1932. Preliminary synopsis by Leonard and H. Y. Moffett published as "Current Definition of Levels in English Usage," *English Journal*, XVI (May, 1927), 345-359.

270. ——————, and H. Y. Moffett. "Current Definition of Levels in English Usage," *English Journal*, XVI (May, 1927), 345-359.

271. Leonard, Vincent T. "A Study of English Composition Assignments in California Junior and Senior High Schools." Unpublished M.A. thesis, San Francisco State College, 1958.

272. Lindquist, Everett F. "The Laboratory Method of Freshman English." Unpublished Ph.D. dissertation, University of Iowa, 1927.

273. Linley, James M. "An Analytical Study of the Creative Process as Revealed by Accounts of Specific Creative Acts." Unpublished Ph.D. dissertation, University of Southern California, 1938.

274. Lis, Anthony S. "Attitudes and Practices of Executives and Secretaries Concerning Disputable Items of English Usage in Secretarial Handbooks." Unpublished Ph.D. dissertation, University of Minnesota, 1961. 285p. UM 61-4603.

275. Littwin, Maxwell F. "An Experimental Investigation of the Effect of Method of Presentation upon the Imaginative Quality of Descriptive Writing among Elementary School Pupils." Unpublished Ph.D. dissertation, New York University, 1934. 315p. Synopsis in *Journal of Experimental Education*, IV (September, 1935), 44-49, and, as "Three Methods of Developing Imagination in Pupil Writing," *English Journal*, XXIV (October, 1935), 654-661.

*276. Loban, Walter. *Language Ability in the Middle Grades of the Elementary School*. Final report to the U.S. Office of Education, Department of Health, Education, and Welfare, 1961. Mimeographed. 145p.

*277. Lokke, Virgil L., and George S. Wykoff. "Double Writing in Freshman Composition: Experiment," *School and Society*, LXVIII (December, 1948), 437-439.

278. Lorge, Irving, and Lorraine Kruglov. "The Relation Between Merit of Written Expression and Intelligence," *Journal of Educational Research*, XLIV (March, 1951), 507-519.

279. ——————, and ——————. "The Relationship Between the Readability of Pupils' Compositions and Their Measured Intelligence," *Journal of Educational Research*, XLIII (February, 1950), 467-474.

280. Lott, Albert J. "Group Composition, Communication, and Consensus: An Investigation According to Newcomb's Theory of Communication." Unpublished Ph.D. dissertation, University of Colorado, 1956.

281. Loughlin, Richard L. "Think Before You Ink," *High Points*, XXXVII (February, 1955), 16-27.

282. Lull, H. G. "The Speaking and Writing Abilities of Intermediate Grade Pupils," *Journal of Educational Research*, XX (June, 1929), 73-77.

283. Lumsen, Florence M. "An Investigation of the Achievement and Growth in Basic English Skills during the First Half of the Senior High School." Unpublished Ed.D. dissertation, George Washington University, 1944. Synopsis in *Journal of Experimental Education*, XIII (September, 1944), 9-14.

284. Lundahl, Walter S., and J. M. Mason. "Essay Testing in Biological Science as a Means for Supplementary Training in Writing Skills," *Science Education*, XL (October, 1956), 261-267.

285. Lupo, Patricia. "A Study of Verb and Pronoun Errors in Original Writing of Ninth-Grade and Twelfth-Grade Pupils of Palmetto High School, Williamston, South Carolina." Unpublished M.A. thesis, Furman University, 1955. 116p.

*286. Lyman, Rollo L. "A Co-operative Experiment in Junior High School Composition," *School Review*, XXXIX (December, 1931), 748-757.

287. Lynch, James J. "The Conference as a Method in the Teaching of English Composition in the Junior-Senior High School." Unpublished Ph.D. dissertation, New York University, 1961. 183p. + ETS STEP test handbooks. UM 61-2558.

288. Macauley, William J. "The Difficulty of Grammar," *British Journal of Educational Psychology*, XVII (November, 1947), 153-162.

289. Mackintosh, Helen K., and Wilhelmina Hill. *How Children Learn to Write.* U.S. Department of Health, Education and Welfare, Office of Education, Bulletin No. 2. Washington: Government Printing Office, 1953.

290. Macmann, Elaine. "A Book Is Born—The Creative Process as Reported by Selected Artists and Observed and Applied in the Writing of a Book for Children." Unpublished Ed.D. dissertation, Columbia University, 1957.

291. Macomber, F. Glenn, and Laurence Siegel. *Final Report of the Experimental Study in Instructional Procedures.* Oxford, Ohio: Miami University, 1960. 96p.

*292. Maize, Ray C. "A Study of Two Methods of Teaching English Composition to Retarded College Freshmen." Unpublished Ph.D. dissertation, Purdue University, 1952. 144p. Synopsis in *Journal of Educational Psychology*, XLV (January, 1954), 22-28, and as "A Writing Laboratory for Retarded Students," *College English*, XVI (October, 1954), 44-48.

293. Malmstrom, Jean. "Linguistic Atlas Findings versus Textbook Pronouncements of Current American Usage," *English Journal*, XLVIII (April, 1959), 191-198.

294. Mann, Mary B. "A Study of the Problem of Quantitative Differentiation of Samples of Written Language." Unpublished Ph.D. dissertation, University of Iowa, 1941. Synopsis in *Psychological Monographs*, LVI, 2 (1944), 41-74.

295. *The Manual Portable Typewriter as an Instructional Tool in the Elementary School Classroom.* Port Chester, New York: Royal McBee Corporation, 1960.

296. Marcus, Fred H. "An Evaluation of the Changes in Certain Communication Skills Effected in College Students by a Freshman Writing Program." Unpublished Ph.D. dissertation, New York University, 1952. 217p. UM 4524.

297. Marquardt, John L. "English Language Instruction in the American High School as Viewed within the Framework of the Adolescent Needs Concept." Unpublished Ed.D. dissertation, Columbia University, 1955.

298. Marshall, Florence R. "A Comparison of Two Methods of Teaching Selected Aspects of English Grammar." Unpublished M.A. thesis, University of New Mexico, 1958.

299. Martin, Clyde I. "An Investigation of the Developmental Interrelationships Among Variables in Children of the First Grade." Unpublished Ph.D. dissertation, University of Texas, 1952.

*300. Matheson, Hugh N. "A Study of Errors Made in Paragraphs Written by Grade 12 Students on the June, 1953, English 40 (Language) University Entrance Examination, British Columbia Department of Education." Unpublished M.A. thesis, University of British Columbia, 1960. 219p.

301. Mathews, Ernest G. "Student English—an All-University Problem," University of Illinois *Faculty Bulletin*, II (April 1, 1949), 1-2.

302. ———, and George Scouffas. "The Freshman's Errors in Composition," *Illinois English Bulletin*, XXXIX (May, 1952), 1-19.

303. ———, Robert P. Larson, and Gibbon Butler. "Experimental Investigation of the Relation between Reading Training and Achievement in College Composition Classes," *Journal of Educational Research*, XXXVIII (March, 1945), 499-505.

304. Mayne, F. Blair. "A Study of Stereotyped Statements: Their Prevalence and Interrelations among a Group of University Students." Unpublished Ph.D. dissertation, New York University, 1942. 298p.

305. McAuliffe, M. F. "A Study of the Writing Vocabularies of Children with IQ's Ranging from Above 120 to Below 86." Unpublished M.A. thesis, Boston University, 1928.

306. McCann, Letitia C. "Writing Vocabularies: A Comparison of the Nature, Extent and Mobility of Child and Adult Writing Needs." Unpublished M.A thesis, University of Texas, 1955. 77p.

*307. McClellan, Jack. "Creative Writing Characteristics of Children." Unpublished Ed.D. dissertation, University of Southern California, 1956. 274p.

308. McCrery, Lester L. "An Experimental Study of Relationships between Writing and Speaking Performance as Measured by College Grades and Student Rating Scales." Unpublished Ph.D. dissertation, University of Southern California, 1950. 212p.

309. McDonald, Catherine B. "Student Preferences in Written-Composition Assignments in the Junior High School," *English Journal*, XXXV (September, 1946), 390-391.

310. McDowell, John G., and Howard R. Anderson. "Testing the Ability of Pupils to Outline," *School Review*, XLVI (January, 1938), 48-56.

311. McEnroe, Kathleen. "The Process of Creative Writing," *Elementary English*, XXXV (March, 1958), 159-162.

312. McGann, Mary. "Diagnostic Testing and Remedial Teaching for Common Errors in Mechanics of English Made by College Freshmen," *Journal of Educational Psychology*, XXXVIII (December, 1947), 499-503.

313. McIntosh, John R. "Learning by Exposure to Wrong Forms in Grammar and Spelling." Contributions to Education, No. 892. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

314. McKee, Paul. "An Adequate Program in the Language Arts," pp. 6-35 in *Teaching Language in the Elementary School*. Forty-Third Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago, 1944.

315. McLaughlin, Iris L. "A Study of the Vocabulary and Sentence Development of Pupils of the Sixth and Seventh Grades of Moore School, Sumter, S.C." Unpublished M.A. thesis, North Carolina College, Durham, 1956.

316. McLellan, George K. "A Comparison of the Usage and Grammar Methods as English Teaching Techniques." Unpublished M.Ed. thesis, University of Texas, 1958.

317. McLendon, Jonathan C. "Are Pupils Learning the Basic Skills of English?" *School Review*, LXI (April, 1953), 226-231.

318. McMechan, Melville Y. "The Relative Effectiveness of Four Procedures for Evaluating Students' Written Themes." Unpublished M.A. thesis, University of British Columbia, 1961. 65p. plus appendices.

*319. Mearns, Hughes. *Creative Power*. New York: Dover Publications, 1959.

*320. ———. *Creative Youth*. New York: Doubleday, Doran, 1925.

321. Mehrens, Harold E. "English Composition in the General College." Unpublished Ph.D. dissertation, University of Southern California, 1940.

322. Mielenz, Mary L. "An Analysis of Background Factors and Their Possible Influence on Placement in Freshman English at the University of Nebraska." Unpublished Ph.D. dissertation, University of Nebraska, 1940.

323. Millar, Janet M. "The Improvement of Spelling as a Tool in Written Expression," pp. 496-502 in *Language Arts in the Elementary School*. Twentieth Yearbook of the Department of Elementary School Principals. Washington: National Education Association, 1941.

324. Miller, George A. *Language and Communication*. New York: McGraw-Hill Book Company, 1951. 298p.

*325. Miller, Joseph W. "An Analysis of Freshman Writing at the Beginning and End of a Year's Work in Composition." Unpublished Ph.D. dissertation, University of Minnesota, 1958. 208p. UM 58-7403.

326. Milligan, John P. "The Effect of Precise and Incidental Teaching of Grammar upon the Written Sentence Structure of Pupils in the Fourth, Fifth, and Sixth Grades." Unpublished Ed.D. dissertation, Rutgers University, 1937.

327. ———. "An Evaluation of Two Methods of Teaching Written Sentence Structure," *Elementary English Review*, XVI (March, 1939), 91-92, 106.

328. ——————. "Learning About Punctuation in the Primary Grades," *Elementary English Review*, XVIII (March, 1941), 96-98.

329. Moe, Sister Mary Leonella. "Teaching Composition in Seventh Grade." Unpublished Ph.D. dissertation, Stanford University, 1956. 168p. UM 56-1441.

330. Monroe, Ilene R. "A Comparative Study of Two Methods of Teaching Fifth Grade English." Unpublished M.Ed. thesis, University of Chattanooga, 1958.

331. Morley, Franklin P. "Creative Behavior in the Elementary School." Unpublished Ed.D. dissertation, Columbia University, 1958.

332. Moody, John L. "An Investigation of the Differences in Language Development between Elementary School Boys and Girls of the Same Chronological Age." Unpublished Ed.D. dissertation, University of Washington, 1956. 92p. UM 18,506.

333. Morrison, R. L., and Philip E. Vernon. "A New Method of Marking English Compositions," *British Journal of Educational Psychology*, XI (June, 1941), 109-119.

334. Mowrer, George E. "A Study of the Effect of the Length of the High School English Class Period on Achievement in English." Unpublished Ed.D. dissertation, University of Missouri, 1956. 136p. UM 56-3342.

335. Moyer, Haverly. "The Effect of Ear Training on Certain Aspects of Children's English Usage." Unpublished Ph.D. dissertation, University of Syracuse, 1951.

336. Neal, Elisa A. "Spelling as Writing." Unpublished Ed.D. dissertation, Columbia University, 1958.

337. Newcomb, William B. "Some Temp Manifestations of the Terminals of English." Unpublished Ph.D. dissertation, University of Wisconsin, 1960.

*338. Newman, Herbert M. "A Study of Range in Expressional Vocabulary." Unpublished Ph.D. dissertation, Columbia University, 1950. 227p. UM 1882.

339. Nisbet, J. D. "English Composition in Secondary School Selection," *British Journal of Educational Psychology*, XXV (February, 1955), 51-54.

340. Noel, Doris I. "A Comparative Study of the Relationship Between the Quality of the Child's Language Usage and the Quality and Types of Language Used in the Home," *Journal of Educational Research*, XLVII (November, 1953), 161-167.

341. Norberg, H. Orville. "Research and the Teaching of Written Expression," *California Journal of Educational Research*, II (March, 1951), 89-91.

*342. Noyes, Edward S. "Essay and Objective Tests in English," *College Board Review*, No. 49 (Winter, 1963), 7-10.

343. O'Donnell, Bernard. *An Analysis of Prose Style to Determine Authorship: A Pilot Study*. Unpublished Special Paper to the Doctoral Committee of the Harvard Graduate School of Education, November, 1962. 63p.

344. Oftedal, Laura. "Picture Writing: A New Tool in Creative Expression," *Elementary School Journal*, IL (September, 1948), 37-46.

345. "Oklahoma City Analysis of Achievement Test Results, 1958-1959." Oklahoma City: Board of Education, [n.d.]. Mimeographed. [concerns teaching by television]

346. Olson, Ove S. "Punctuation Situations in Ninth Grade Writing." Unpublished M.A. thesis, University of Iowa, 1926.

347. Oseas, Leonard. "A Study of Some Language Correlates of Passivity in Personality." Unpublished Ph.D. dissertation, Northwestern University, 1954. 95p. UM 9266.

*348. Parker, John P. "Some Organizational Variables and Their Effect upon Comprehension," *Journal of Communication*, XII (March, 1962), 27-32.

349. Parker, Katherine, *et al.* "The Construction and Evaluation of Five Patterned Series of Lessons for Teaching Five Specific Writing Skills." Unpublished M.A. thesis, Boston University, 1959.

350. Parler, Nettie P. "Writing Skills of Prospective English Teachers." Unpublished Ph.D. dissertation, New York University, 1952. 125p. UM 4526.

351. Patrick, Catharine. "Creative Thought in Poets." *Archives of Psychology*, No. 178 (April, 1935), 74p.

352. Pearson, Richard, and Earle G. Eley. "Should the General Composition Test Be Continued?" *College Board Review*, No. 25 (Winter, 1955), 2-13.

353. Pease, Marion O. "English Errors in Social Studies Notebooks," *Elementary English Review*, XVIII (February, 1941), 47-51.

354. Peel, E. A., and H. G. Armstrong. "The Predictive Power of the English Composition in the 11+ Examination," *British Journal of Educational Psychology*, XXVI (November, 1956), 163-171.

355. Peery, William. "Freshman English and the Enrollment Bulge," *Association of American Colleges Bulletin*, XLI (December, 1955), 575-583.

356. Pendleton, Paul E. "The Interdepartmental Teaching of English in College: Its Extent, Its Methods, and Its Possibilities." Unpublished Ph.D. dissertation, The Ohio State University, 1941.

357. Penfold, D. M. Edwards. "Essay Marking Experiments: Shorter and Longer Essays," *British Journal of Educational Psychology*, XXVI (June, 1956), 128-136.

358. Perrin, Porter G. "Commensurate Grading of English Compositions," *School and Society*, XXXVII (March 25, 1933), 386-387.

359. Phillips, Phyllis J. "The Relationship between a Knowledge of Correct English Usage and Composition Writing, as Demonstrated by College Freshmen." Unpublished Ed.D. dissertation, Colorado State College of Education, 1955. 293p. UM 20,336.

360. Piaget, Jean. *The Language and Thought of the Child.* 3d ed., rev. and enl. Tr. Marjorie Warden. New York: Humanities Press, 1959. 288p.

361. Pidgeon, D.A., and Alfred Yates. "Experimental Inquiries into the Use of Essay-Type English Papers," *British Journal of Educational Psychology*, XXVII (February, 1957), 37-47.

362. Platt, Frank J. "Measuring Improvement in Language Essentials," *School Review*, XXVIII (October, 1920), 611-614.

363. Platt, James H. "A Study of Personality Factors Common to Speaking, Listening, Writing and Reading." Unpublished Ph.D. dissertation, University of Denver, 1954. 267p.

364. Polychrones, James Z. "Children's Usage of Punctuation and Capitalization in the Intermediate Grades." Unpublished Ph.D. dissertation, Northwestern University, 1958. 287p. UM 58-5776.

365. Pollock, Thomas C. "Spelling Report," *College English*, XVI (November, 1954), 102-109.

366. Pond, Frederick L. "Influence of the Study of Latin on Word Knowledge," *School Review*, XLVI (October, 1938), 611-618.

367. Pooley, Robert C. "Language Arts Survey in Wisconsin Elementary Schools," *Elementary English Review*, XXIII (January, 1946), 8-14.

368. ———. *Teaching English Usage.* National Council of Teachers of English, English Monograph No. 16. New York: Appleton-Century-Crofts, 1946. 265p.

369. ———, and Robert D. Williams. *The Teaching of English in Wisconsin.* Madison: University of Wisconsin Press, 1948. 196p.

370. Powell, Raymond L. "Valid Testing and Diagnosis in the Mechanics of Ninth Grade English Composition," University of Iowa *Studies in Education*, IX (January, 1934), 95-130.

371. Pressey, S. L., and Pere Campbell. "The Causes of Children's Errors in Capitalization: A Psychological Analysis," *English Journal*, XXII (March, 1933), 197-201.

372. Price, Joseph S. C. "Predicting Freshman Achievement: A Follow-up Study." Unpublished Ph.D. dissertation, Harvard University, 1950.

373. Quant, Leslie. "Factors Affecting the Legibility of Handwriting." Unpublished Ph.D. dissertation, University of Chicago, 1945.

374. Ransom, Grace. "Remedial Methods in English Composition," *English Journal* (High School Edition), XXII (November, 1933), 749-754.

375. Remmers, H. H., *et al.* "Concerning Freshman Composition Tangibles and Intangibles of Achievement," Purdue University *Studies in Higher Education*, XXXV, 2 (1934). 34p.

376. Rhoades, Inez B. "Personality Trends of Speakers and Writers in Junior High School." Unpublished M.A. thesis, University of Denver, 1942.

377. Richardson, Henry M. "Where Do We Stand in Grammar Instruction?" Unpublished Ed.D. dissertation, Columbia University, 1953. 171p.

378. Rider, Maurice L. "Advanced Composition for Students in Engineering at Ohio State University: Evaluation and Proposals." Unpublished Ph.D. dissertation, The Ohio State University, 1950.

*379. Rinsland, Henry D. *A Basic Vocabulary of Elementary School Children.* New York: Macmillan Company, 1945. 636p.

380. Ritter, Deckard. "Five Years of Directing Extra-Curriculum Creative Writing in a College." Unpublished Ed.D. dissertation, New York University, 1935.

381. Robbins, Wilhelmina K. "Spelling Errors in the Writing of Second and Third Grade Children." Unpublished M.A. thesis, University of Iowa, 1948. 174p.

382. Robinson, N. "The Relation Between Knowledge of English Grammar and Ability in English Composition." Unpublished M.Ed. thesis, Manchester University, 1959. 110p.

383. Rodgers, G. S. "A Critical Study of the Grammatical Errors of Junior High School Pupils," *Journal of Applied Psychology*, XVI, 4 (1932), 421-426.

384. Rollstin, Donald H. "A Study of the Handwriting of College Freshmen." Unpublished M.A. thesis, University of Iowa, 1949. 40p.

385. Rosenthal, Fred. "Some Relationships between Sociometric Position and Language Structure of Young Children," *Journal of Educational Psychology*, XLVIII (December, 1957), 483-497.

386. Rothenberg, Julius G. "English Errors of Slow Learners," *English Journal*, XXXII (December, 1943), 551-556.

*387. Rounds, Robert W. "Teaching Creative Writing in the Junior High School." Unpublished Ed.D. dissertation, New York University, 1942.

388. Runkel, Philip J. "Cognitive Facilitation of Communication Effects: An Empirical Study." Unpublished Ph.D. dissertation, University of Michigan, 1956. 160p. UM 56-5357.

389. Russell, David H. *Characteristics of Good and Poor Spellers.* Contributions to Education, No. 727. New York: Bureau of Publications, Teachers College, Columbia University, 1937.

390. Salisbury, Rachel. "A Study of the Effects of Training in Logical Organization as a Method of Improving Skill in Study." Unpublished Ph.D. dissertation, University of Wisconsin, 1934. Synopsis in *Journal of Educational Research*, XXVIII (December, 1934), 241-254.

391. Salt, Edna. "Language Instruction: Formal vs. Functional," *Childhood Education*, XXVI (November, 1949), 124-127.

392. Sauer, Edwin H. *Contract Correcting: The Use of Lay Readers in the High School Composition Program.* Report to the Cooperative Research Program of the United States Department of Health, Education, and Welfare. Rev. ed., 1962. 59p.

393. Schimek, Jean G. "Creative Originality: Its Evaluation by the Use of Free-Expression Tests." Unpublished Ph.D. dissertation, University of California, Berkeley, 1954. 106p.

394. Schmidt, Sister M. Aurelian. "An Experimental Comparison of Socialized and Teacher Correction Methods in English Composition." Unpublished M.A. thesis, Marquette University, 1960.

395. Schonell, F. J. *Backwardness in the Basic Subjects.* London: Oliver and Boyd, 1949. 566p.

396. Scrafford, Ralph. "A Study of the Desirability of Correlating English Composition with Social Studies." Unpublished Ed.D. dissertation, (Pennsylvania State College) The Pennsylvania State University, 1943. 214p. UM 653.

397. "Second Report of the National Experiment of Television Teaching in Large Classes: 1958-1959." Philadelphia: Philadelphia Public Schools, [n.d.]. Mimeographed.

*398. Seegers, J. C. "Form of Discourse and Sentence Structure," *Elementary English Review*, X (March, 1933), 51-54.

399. Seibert, W. F. "An Evaluation of Televised Instruction in College English Composition." Lafayette, Indiana: Purdue University, 1958. Mimeographed.

400. Seidel, Aimme A. "Case Studies of Thirty Maladjusted Pupils in Tenth A and Eleventh B English." Unpublished M.A. thesis, University of Pittsburgh, 1934.

401. Schrader, Helen W. "A Linguistic Approach to the Study of Rhetorical Style." Unpublished Ph.D. dissertation, Northwestern University, 1949.

402. Serra, Mary C. "Influence of Reading and Experience-Writing on Spelling: A Case Study," *Elementary School Journal*, LIV (April, 1954), 456-463.

403. Shaffer, Robert H. "The Effect of English Deficiency Upon a Student's Adjustment in College," *Bulletin of the School of Education*, Indiana University, XXIV, 1 (January, 1948), 1-34.

404. Sherman, Dorothy H. "An Objective Study of the Ability of Junior High School Pupils to Use Descriptive Language." Unpublished M.A. thesis, University of Denver, 1938.

405. Sherwood, John C. "Final Report: Composition Experiment: University of Oregon." Mimeographed, undated. 21p. Synopsis in *College Composition and Communication*, IX (February, 1958), 5-9.

406. Silverthorn, James. "The Basic Vocabulary of Written Business Communications." Unpublished Ph.D. dissertation, Indiana University, 1955.

407. Simpson, Loren E. "An Evaluation of Usage Errors Made by 1135 Kansas High School Seniors." Unpublished M.A. thesis, University of Kansas, 1957.

408. Sims, Verner M. "Improving the Measuring Qualities of an Essay Examination," *Journal of Educational Research*, XXVII (September, 1933), 20-31.

409. Singleton, Carlton M. "Imagery: Measurement and Analysis." Unpublished Ed.D. dissertation, Boston University. 1955.

410. Skelton, Robert B. "High-School Foreign Language Study and Freshman Performance," *School and Society*, LXXXV (June 8, 1957), 203-205.

411. Skuba, Michael. "An Analysis of English Errors and Difficulties among Grade Ten Students in the Smoky Lake Division." Unpublished M.Ed. thesis, University of Alberta, 1955. 106p. Synopsis in *Alberta Journal of Educational Research*, December, 1955.

412. Sloan, Geraldine. "An Application of Structural Grammar in the Teaching of Sentence Construction to Sub-Freshman College Students." Unpublished M.A. thesis, Chico State College, 1957.

*413. Smith, Dora V. *Class Size in High School English: Methods and Results.* Minneapolis: University of Minnesota Press, 1931. 309p.

414. ——————. *Evaluating Instruction in Secondary School English.* A Report of a Division of the New York Regents' Inquiry into the Character and Cost of Public Education in New York State. English Monograph No. 11. Chicago: National Council of Teachers of English, 1941. 273p.

415. ——————. "Growth in Language Power as Related to Child Development," pp. 52-97 in Forty-Third Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1944.

416. ——————. "Implications of the New York Regents' Inquiry for Teaching English," *English Journal* (High School Edition), XXVIII (March, 1939), 177-192.

417. ——————. "Problems of Class Size and the Efficiency of Instruction in English," *English Journal*, XIX (November, 1930), 724-726.

418. Smith, Ethel E. "Procedures for Encouraging Creative Writing in the Elementary School." Unpublished Ph.D. dissertation, Northwestern University, 1944.

419. Smith, Geraldine M. "A Comparative Study of the Oral and Written Methods

in the Teaching of English Usage." Unpublished M.A. thesis, University of Iowa, 1959.

420. Smith, Henry L., Jr. *Linguistic Science and the Teaching of English.* Cambridge: Harvard University Press, 1956.

421. ————, and Merrill T. Eaton. "A Study of the English Usage, Spelling, and Vocabulary of 251 Graduate Students at Indiana University," *Bulletin of the School of Education,* Indiana University, XVI, 5 (October, 1940), 1-71.

422. Smith, James A. "Creative Teaching in Elementary Schools: Its Basis in Research with Illustrations of Practice in Selected Areas." Unpublished Ph.D. dissertation, University of Syracuse, 1956.

423. Smith, M. E. "Investigation of the Development of the Sentence and the Extent of Vocabulary in Young Children." University of Iowa *Studies in Child Welfare,* III, 5 (1926). 92p.

424. Solterer, Josef. "Retesting Freshman English," *School and Society,* LIV (October 11, 1941), 305.

425. Spache, George. "Characteristic Errors of Good and Poor Spellers," *Journal of Educational Research,* XXXIV (November, 1940). 182-189.

426. Spear, Josephine. "A Comparison of Two Methods of Teaching Grammatical Usage." Unpublished Ed.D. dissertation, Indiana University, 1951. 362p.

427. Spitzer, Herbert F. "Class Size and Pupil Advancement in Elementary Schools," *Elementary School Journal,* LV (October, 1954), 82-86.

428. Stageberg, Norman C. "Some Structural Ambiguities," *English Journal,* XLVIII (November, 1958), 479-486.

*429. Stalnaker, John M. "The Construction and Results of a Twelve-Hour Test in English Composition," *School and Society,* XXXIX (February 17, 1934), 218-224.

*430. ————. "Essay and Objective Writing Tests," *English Journal* (College Edition), XXII (March, 1933), 217-222.

431. ————. "Objective Placement Tests in English Composition," *English Journal* (College Edition), XXII (May, 1933), 395-403.

432. ————. "Recognition of Errors versus Their Correction," *English Journal* (High School Edition), XXII (June, 1933), 493-494.

433. ————. "Sex Differences in the Ability to Write," *School and Society,* LIV (December 6, 1941), 532-535.

*434. Starring, Robert W. "A Study of Ratings of Comprehensive Examination Themes When Certain Elements Are Weakened." Unpublished Ed.D. dissertation, Michigan State College, 1952. 109p. UM 4327.

435. Stecklein, John E. "Equivalence of Word Fluency Factor Patterns." Unpublished Ph.D. dissertation, University of Wisconsin, 1953.

436. Stewart, James R. "The Effect of Diagramming on Certain Skills in English Composition." Unpublished Ph.D. dissertation, University of Iowa, 1942. Synopsis in *Journal of Experimental Education,* XI (September, 1942), 1-8.

437. Stillberger, Evelyn C. "Individual Differences of Achievement in Spelling." Unpublished M.A. thesis, University of Iowa, 1950. 135p.

438. Stormzand, Michael, and M. V. O'Shea. *How Much English Grammar?* Baltimore: Warwick and York, 1924.

439. Strickland, Ruth G. "The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children," *Bulletin of the School of Education,* Indiana University, XXXVIII, 4 (July, 1962), 1-131.

440. Suggs, Lena R. "Structural Grammar versus Traditional Grammar in Influencing Writing," *English Journal,* L (March, 1961), 174-178.

441. Swenson, Esther J., and Charles G. Caldwell. "The Content of Children's Letters," *Elementary School Journal,* IL (November, 1948), 149-159.

442. ————, and ————. "The Process of Communication in Children's Letters," *Elementary School Journal,* IL (October, 1948), 79-88.

443. ————, and ————. "Spelling in Children's Letters," *Elementary School Journal,* IL (December, 1948), 224-235.

•444. Swineford, Francis. "Reliability of an Interlinear Test of Writing Ability." *School and Society*, LXXXI (January 22, 1955), 25-27.

445 Symonds, Percival M. "Practice versus Grammar in the Learning of Correct English Usage," *Journal of Educational Psychology*, XXII (February, 1931), 81-95.

446. ———, and Baldwin Lee. "Studies in the Learning of English Expression: I—Punctuation; II—Capitalization," *Teachers College Record*, XXX (February and April, 1929), 461-480 and 686-692.

447. ———, and Helen F. Daringer. "Studies in the Learning of English Expression: IV—Sentence Structure," *Teachers College Record*, XXXII (October, 1930), 50-64.

448. Taylor, Calvin W. "A Factorial Study of Fluency in Writing." Unpublished Ph.D. dissertation, University of Chicago, 1947.

449. *Teaching by Television*. New York: The Ford Foundation and the Fund for the Advancement of Education, 1959.

450. Templin, Mildred C. "A Comparison of the Spelling Achievement of Normal and Defective Hearing Subjects," *Journal of Educational Psychology*, XXXIX (October, 1948), 337-346.

451. Thomas, J. E. "The Elimination of Technical Errors in Written Compositions through Formal Drill." University of Iowa *Studies in Education*, VIII, 2 (1932).

452. Thomas, Joseph M., and Nicholas Fattu. "A Comparison of Laboratory and Nonlaboratory Instruction in Freshman English," pp. 85-90 in *Studies in Higher Education*. Biennial Report of the Committee on Education Research, 1938-1940. Minneapolis: University of Minnesota, 1941.

453. Thompson, Stanley I. "Integration of Fifth Grade Spelling and Handwriting," *Elementary School Journal*, XLII (January, 1942), 347-357.

454. Thompson, Wayne N. "A Study of the Grading Practices of Thirty-One Instructors in Freshman English," *Journal of Educational Research*, IL (September, 1955), 65-68.

•455. Thorndike, Edward L., *et al.* "An Inventory of English Constructions with Measures of Their Importance," *Teachers College Record*, XXVIII (February, 1927), 580-610.

•456. ———, and Irving Lorge. *The Teacher's Word Book of 30,000 Words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

457. Tondow, Murray. "A Study of Oral and Written Language of Children in the Fourth, Fifth, and Sixth Grades." Unpublished Ed.D. dissertation, University of Southern California, 1953. 267p.

458. Toporowski, Dolores E. "A Study of the Opinions and Actions of Elementary School Teachers and Pupils Regarding the Correcting of Papers." Unpublished M.A. thesis, Boston University, 1958.

459. Torgerson, Warren S., and Bert F. Green, Jr. "The Factor Analysis of Subject-Matter Experts," *Journal of Educational Psychology*, XLIII (October, 1952), 354-363.

460. Tressler, Jacob C. "The Efficiency of Student Correction of Compositions," *English Journal*, I (September, 1912), 405-411.

461. ———. "Salvaging from the English Scrap Heap," *English Journal*, VIII (September, 1919), 412-418.

462. Triggs, Francis O. "The Role of Tests in the Diagnosis and Correction of Spelling Deficiencies of College Students," *Educational and Psychological Measurement*, V (Spring, 1945), 59-70.

•463. Van Bruggen, John A. "Factors Affecting Regularity of the Flow of Words During Written Composition." Unpublished Ph.D. dissertation, University of Chicago, 1945. 118p. Synopsis in *Journal of Experimental Education*, XV (December, 1946), 133-155.

464. Virtue, John B. "The Proficiency Examination in English Composition at the University of Kansas," *College English*, IX (January, 1948), 199-203.

465. "Visual Stimulus to Compositions: Experimental Film," *Times Educational Supplement*, June 14, 1947, p. 291.

*466. Vordenberg, Wesley P. "A Comparison of Errors Made by College Freshmen on the Cooperative English Test, Form PM, and in Themes." Unpublished Ph.D. dissertation, Indiana University, 1950. Synopsis, as "How Valid Are Objective English Tests?" *English Journal*, XLI (October, 1952), 428-429.

467. Wagner, Carl E. "Experimental Evaluation of a Method of Teaching English Composition." Unpublished Ph.D. dissertation, University of Kansas, 1956.

468. Walsh, William E. "One Method of Removing Some of the Blockages to a Broader Communication of Validity in Certain Areas of Creative Writing." Unpublished M.A. thesis, University of Denver, 1947.

469. Warbois, G. M. "Language Development of Children in Two Types of Rural Environments," *Journal of Experimental Education*, X (December, 1941), 150-152.

470. Ward, Louis R. "A Comparison of Two Methods of Teaching Writing in a Course in Communication Skills." Unpublished Ph.D. dissertation, Purdue University, 1959. 99p. UM 59-6523.

471. Warner, Paul C., and W. S. Guiler. "Individual vs. Group Instruction in Grammatical Usage," *Journal of Educational Psychology*, XXIV (February, 1933), 140-151.

472. Wason, P. C. "The Effect of Compressing Information on Its Retention." Unpublished Ph.D. dissertation, University of London, 1957. 195p. Synopsis, as "The Retention of Material Presented Through Précis," *Journal of Communication*, XII (March, 1962), 36-43.

473. Watts, A. F. Chapter V, "Learning to Write English," pp. 119-144 in *The Language and Mental Development of Children*. Boston: D. C. Heath and Company, 1944.

474. Weaver, Ella H. "An Approach to Language Behavior from the Point of View of General Semantics." Unpublished M.A. thesis, Northwestern University, 1949.

475. Weber, John L. "An Analysis of Class Load and Frequency of Writing Assignments in English Classes of Southwestern Michigan High Schools." Unpublished M.A. thesis, Western Michigan University, 1958.

476. Weigand, Elsie F. "The Effectiveness of Composition of Formalized Grammar Teaching versus Work in Correct Language Usage." Unpublished Ed.D. dissertation, New York University, 1944.

477. Werner, Oscar H. "The Influence of the Study of Modern Foreign Languages on the Development of Desirable Abilities in English," pp. 99-145 in *Studies in Modern Language Teaching*. Publications of the American and Canadian Committees on Modern Languages, Vol. XVIII. New York: Macmillan Company, 1930.

478. Wheeler, Arville. "Study to Determine the Errors That Appear in Written Work of Rural and Urban Pupils in Certain School Systems in Kentucky." Unpublished Ph.D. dissertation, Cornell University, 1940. Synopsis in *Journal of Experimental Education*, VIII (June, 1940), 385-398.

479. Whitaker, Virgil K. "Report on a Study of English Composition Undertaken by the Department of English in Cooperation with the Stanford Study of Undergraduate Education." Mimeographed. 14p.

480. Wild, Mollie K. "Creative Writing and Achievement in Solving Developmental Tasks in the Elementary School." Unpublished Ed.D. dissertation, The Pennsylvania State University, 1957. 57p. UM 24,048.

481. Wilhelms, Fred T. "Individualized Remedial Instruction in English Usage at the High School Level." Unpublished Ph.D. dissertation, University of Nebraska, 1940.

*482. Willing, Matthew H. *Valid Diagnosis in High School Composition*. Contributions to Education, No. 230. New York: Bureau of Publications, Teachers College, Columbia University, 1926. 64p. Synopsis in *Journal of Educational Research*, XIII (February, 1926), 77-89.

483. Winter, Ralph D. "English Function Words and Content Words: A Quantitative Investigation." Unpublished Ph.D. dissertation, Cornell University, 1953. 341p. UM 8231.

484. Wise, Harold L. "An Analysis of the Written Language of Children in Grades One, Two, and Three." Unpublished M.A. thesis, University of Iowa, 1959.

*485. Wiseman, Stephen. "The Marking of English Composition in Grammar School Selection," *British Journal of Educational Psychology*, XIX (November, 1949), 200-209. Synopsis in *Times Educational Supplement*, April 9, 1949, p. 231.

486. ———. "Reliability and Validity," *British Journal of Educational Psychology*, XXVI (November, 1956), 172-179.

*487. ———, and Jack Wrigley. "Essay-Reliability: The Effect of Choice of Essay-Title," *Educational and Psychological Measurement*, XVIII (Spring, 1958), 129-138.

488. Wiswall, Zilla E. "Improvement in Sentence Structure," *English Journal*, XVI, (September, 1927), 518-527.

489. Witty, Paul. "Creative Writing in Elementary School," *Schoolman's Week Proceedings*, XXIX (1942), 127-135.

490. ———. *School Children and Television*. New York: Television Information Office, 1960. 26p.

491. ———, and Mabel Fry. "The Vocabulary Content of Compositions Written by College Students," *Journal of Educational Research*, XIX (February, 1929), 135-139.

492. ———, and Roberta L. Green. "Composition Errors of College Students," *English Journal* (College Edition), XIX (May, 1930), 388-393.

493. ——— and Lou L. LaBrant. "Vocabulary and Reading," *School and Society*, XXXI (February 22, 1930), 268-272.

494. ———, and William Martin. "An Analysis of Children's Compositions Written in Response to a Film," *Elementary English*, XXXIV (March, 1957), 158-163.

495. Wixted, William G., and Philip R. V. Curoe. "How Well Do College Seniors Write?" *School and Society*, LIV (November 29, 1941), 505-508.

496. Wolfson, Rose. "A Study in Handwriting Analysis." Unpublished Ph.D. dissertation, Columbia University, 1949.

497. Womack, William T. "A Study of Teachers' Attitudes toward Debatable Items of English Usage." Unpublished Ph.D. dissertation, Columbia University, 1957. 198p. UM 58-1368.

*498. Wood, Ben D., and Frank N. Freeman. *An Experimental Study of the Educational Influences of the Typewriter in the Elementary School Classroom*. New York: Macmillan Company, 1932.

*499. Wormsbecker, J. H., Jr. "A Comparative Study of Three Methods of Grading Compositions." Unpublished M.A. thesis, University of British Columbia, 1955. 44p.

500. Wykoff, George S. "Freshman Composition without Class Attendance: An Experiment," *School and Society*, LV (March, 1942), 276-280.

501. ———. "The Relation of a Knowledge of Grammar and Punctuation to Writing," *Educational Administration and Supervision*, XXXI (October, 1945), 385-393.

502. Yalden, Maxwell F. "Language and Cognition: An Examination of the Hypothesis that Language Influences Habitual Perception and Thought." Unpublished Ph.D. dissertation, University of Michigan, 1956. 327p. UM 21,376.

503. Young, Florence. "An Analysis of Certain Variables in a Developmental Study of Language." Unpublished Ph.D. dissertation, George Peabody College, 1939.

504. Zedler, Empress Y. "Effect of Phonic Training on Speech Sound Discrimination and Spelling Performance," *Journal of Speech and Hearing Disorders*, XXI (June, 1956), 245-250.